STIC-Biotech/ChemLib 60001

From: Ramirez, Delia
Sent: Friday, February 08, 2002 7:08 PM
To: STIC-Biotech/ChemLib
Subject: case 09/606129

Hi,

I would like to request the following searches (09/606129 Maines et al.)

1. a standard search of seq id 1, 3, 18, 19, 34, 35 in the protein databases (commercial and interference)
2. an oligo search of seq id 18, 19, 34, 35 in the protein databases (commercial)

Thank you,

Delia M. Ramirez, Ph.D.
Patent Examiner - Art Unit 1652
USPTO
1911 S. Clark Street, Crystal Mall 1, 10D04, Mail room 10C01
Arlington, VA 22202
(703) 306-0288
delia.ramirez@uspto.gov

	TYPE OF SEARCH:	VENDOR/COST(where applic.)
Searcher:	NA Sequences:	STN:
Phone:	AA Sequences: 10	DIALOG:
Location:	Structures:	Questel/Orbit:
Date Picked Up: 2/11/02	Bibliographic:	DRLink:
Date Completed: 2/12/02	Litigation:	Lexis/Nexis:
Searcher Prep/Review:	Full text:	Sequence Sys.: OK
Clerical:	Patent Family:	WWW/Internet:
Online time:	Other:	Other (specify):

Pending Nucleic Acid and/or Pending Amino Acid database searches now generate two sets of results. These databases were split into to two parts to reduce the time needed to update the databases daily. The split freed up more machine time for processing searches.

Searches run against the Nucleic Acid Pending database produce two sets of results, with the extensions, **.rnpm** and **.rnpn**
Searches run against the Amino Acid Pending database produce two sets of results, with the extensions, **.rapm** and **.rapn**

The Pending database search results should not be left in the case because they contain data that is confidential.

GenCore version 4.5

OM protein - protein search, using sw model

Run on: February 12, 2002, 12:03:24 ; Search time 94.82 Seconds
(without alignments)
10.798 Million cell updates/sec

Title: US-09-606-129A-18
Perfect score: 41
Sequence: 1 KKRIMHC 7

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_17:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	41	100.0	599	10	Q9FHV3	Q9fhv3 arabidopsis
2	41	100.0	633	10	O49511	O49511 arabidopsis
3	38	92.7	295	11	Q9CY64	Q9cy64 mus musculu
4	38	92.7	296	4	Q9BRW8	Q9brw8 homo sapien
5	38	92.7	303	11	Q9DD21	Q9dd21 mus musculu
6	37	90.2	106	5	Q9XZM5	Q9xzm5 leishmania
7	36	87.8	512	10	Q9LML0	Q9lml0 arabidopsis
8	35	85.4	508	11	Q9D6S2	Q9d6s2 mus musculu
9	34	82.9	279	11	Q9QXV3	Q9qxv3 mus musculu
10	34	82.9	401	10	Q9M149	Q9m149 arabidopsis
11	34	82.9	431	10	O04613	O04613 arabidopsis
12	33	80.5	189	2	Q03949	Q03949 anabaena sp
13	33	80.5	267	2	Q9AKS0	Q9aks0 pseudomonas
14	33	80.5	495	10	Q9ST63	Q9st63 solanum tub
15	33	80.5	508	10	O80874	O80874 arabidopsis
16	32	78.0	102	10	Q9S738	Q9s738 lycopersico
17	32	78.0	104	12	Q9Q553	Q9q553 human immun
18	32	78.0	108	2	P75909	P75909 escherichia
19	32	78.0	111	4	Q9H007	Q9h007 homo sapien
20	32	78.0	142	4	Q9BTS7	Q9bts7 homo sapien
21	32	78.0	233	4	Q9UIJ2	Q9uij2 homo sapien
22	32	78.0	235	4	Q9P0U6	Q9p0u6 homo sapien
23	32	78.0	279	4	O43658	O43658 homo sapien
24	32	78.0	279	4	Q9UK52	Q9uk52 homo sapien
25	32	78.0	279	4	Q9HD98	Q9hd98 homo sapien
26	32	78.0	283	4	Q9HD99	Q9hd99 homo sapien
27	32	78.0	294	4	O00532	O00532 homo sapien
28	32	78.0	338	3	Q9C1L3	Q9c1l3 neurospora
29	32	78.0	377	2	Q9KSC1	Q9ksc1 vibrio chol
30	32	78.0	422	4	Q9UIJ4	Q9uij4 homo sapien
31	32	78.0	422	4	Q9UK53	Q9uk53 homo sapien
32	32	78.0	455	5	Q9VJ28	Q9vj28 drosophila
33	32	78.0	497	10	Q9C8V1	Q9c8v1 arabidopsis
34	32	78.0	516	5	Q21592	Q21592 caenorhabdi
35	32	78.0	571	10	O24022	O24022 lycopersico
36	32	78.0	601	4	Q9BRB5	Q9brb5 homo sapien
37	32	78.0	978	5	Q9U6W5	Q9u6w5 caenorhabdi
38	31	75.6	109	5	Q9VZN4	Q9vzn4 drosophila
39	31	75.6	137	4	Q9NSR1	Q9nsr1 homo sapien
40	31	75.6	152	5	O44797	O44797 caenorhabdi
41	31	75.6	279	2	Q9AJ03	Q9aj03 staphylococ
42	31	75.6	321	10	Q9LZD9	Q9lzd9 arabidopsis
43	31	75.6	398	5	Q9VT97	Q9vt97 drosophila
44	31	75.6	446	2	Q9CKP7	Q9ckp7 pasteurella
45	31	75.6	476	4	Q9UFK0	Q9ufk0 homo sapien

ALIGNMENTS

RESULT 1
Q9FHV3
ID Q9FHV3 PRELIMINARY; PRT; 599 AA.
AC Q9FHV3;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE MUTATOR-LIKE TRANSPOSASE-LIKE PROTEIN.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RX MEDLINE=99397451; PubMed=10470850;
RA Kaneko T., Katoh T., Sato S., Nakamura Y., Asamizu E., Kotani H.,
RA Miyajima N., Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. IX.
RT Sequence features of the regions of 1,011,550 bp covered by seventeen
RT P1 and TAC clones.";
RL DNA Res. 6:183-195(1999).
DR EMBL; AB017068; BAB11366.1; -.
SQ SEQUENCE 599 AA; 69407 MW; 36BAEE2F2AB2D717 CRC64;

Query Match 100.0%; Score 41; DB 10; Length 599;
Best Local Similarity 100.0%; Pred. No. 0.73;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 KKRIMHC 7
          |||||||
Db    563 KKRIMHC 569
```

RESULT 2
O49511
ID O49511 PRELIMINARY; PRT; 633 AA.
AC O49511;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)

```
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   MUDR TRANSPOSABLE ELEMENT - LIKE PROTEIN (MUDR TRANSPOSABLE ELEMENT-
DE   LIKE PROTEIN).
GN   F28J12.70 OR AT4G18410.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Bevan M., Hilbert H., Braun M., Holzer E., Brandt A., Duesterhoeft A.,
RA   Bancroft I., Mewes H.W., Mayer K., Schueller C.;
RL   Submitted (FEB-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Hilbert H., Braun M., Holzer E., Brandt A., Duesterhoeft A.,
RA   Mewes H.W., Lemcke K., Mayer K.F.X.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   EU Arabidopsis sequencing project;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL021710; CAA16721.1; -.
DR   EMBL; AL161548; CAB78843.1; -.
DR   Mendel; 27564; Arath;3233;27564.
SQ   SEQUENCE   633 AA;  72930 MW;  3FC298BF2218C623 CRC64;


  Query Match             100.0%;  Score 41;  DB 10;  Length 633;
  Best Local Similarity   100.0%;  Pred. No. 0.76;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 KKRIMHC 7
           |||||||
Db     591 KKRIMHC 597


RESULT   3
Q9CY64
ID   Q9CY64      PRELIMINARY;      PRT;   295 AA.
AC   Q9CY64;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   2500001N03RIK PROTEIN.
GN   2500001N03RIK.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=EMBRYONIC LIVER;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
DR   EMBL; AK010847; BAB27219.1; -.
DR   MGD; MGI:1917355; 2500001N03Rik.
DR   InterPro; IPR000683; GFO_IDH_MocA.
DR   Pfam; PF01408; GFO_IDH_MocA; 1.
SQ   SEQUENCE   295 AA;  33524 MW;  F2E1682BD77032A4 CRC64;


  Query Match             92.7%;  Score 38;  DB 11;  Length 295;
  Best Local Similarity   85.7%;  Pred. No. 1.8;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 KKRIMHC 7
           ||||:||
Db     274 KKRILHC 280


RESULT   4
Q9BRW8
ID   Q9BRW8      PRELIMINARY;      PRT;   296 AA.
AC   Q9BRW8;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   SIMILAR TO BILIVERDIN REDUCTASE A.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RA   Strausberg R.;
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC005902; AAH05902.1; -.
SQ   SEQUENCE   296 AA;  33432 MW;  A389AFDDB6ADBABF CRC64;


  Query Match             92.7%;  Score 38;  DB 4;  Length 296;
  Best Local Similarity   85.7%;  Pred. No. 1.8;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 KKRIMHC 7
           ||||:||
Db     275 KKRILHC 281


RESULT   5
Q9DD21
ID   Q9DD21      PRELIMINARY;      PRT;   303 AA.
AC   Q9DD21;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   0610006A11RIK PROTEIN.
GN   0610006A11RIK.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=KIDNEY;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
```

```
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
DR   EMBL; AK002231; BAB21950.1; -.
DR   MGD; MGI:1915580; 0610006A11Rik.
DR   InterPro; IPR000683; GFO_IDH_MocA.
DR   Pfam; PF01408; GFO_IDH_MocA; 1.
SQ   SEQUENCE   303 AA;  34491 MW;  52D8A3B02EE956EB CRC64;


  Query Match             92.7%;  Score 38;  DB 11;  Length 303;
  Best Local Similarity   85.7%;  Pred. No. 1.8;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 KKRIMHC 7
           ||||:||
Db     274 KKRILHC 280


RESULT   6
Q9XZM5
ID   Q9XZM5      PRELIMINARY;      PRT;   106 AA.
AC   Q9XZM5;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   RIBOSOMAL PROTEIN L44.
GN   RPL44.
OS   Leishmania amazonensis.
OC   Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX   NCBI_TaxID=5659;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LV78;
RA   Porter-Kelley J., Chaudhuri G.;
RT   "Cloning and characterization of Leishmania ribosomal protein L44.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: TO THE L44E FAMILY OF RIBOSOMAL PROTEINS.
DR   EMBL; AF148853; AAD31928.2; -.
DR   InterPro; IPR000552; Ribosomal_L44E.
DR   Pfam; PF00935; Ribosomal_L44; 1.
DR   ProDom; PD002841; Ribosomal_L44E; 1.
DR   PROSITE; PS01172; RIBOSOMAL_L44E; 1.
KW   Ribosomal protein.
SQ   SEQUENCE   106 AA;  12283 MW;  F30A3AB2047B0334 CRC64;


  Query Match             90.2%;  Score 37;  DB 5;  Length 106;
  Best Local Similarity   71.4%;  Pred. No. 1.2;
  Matches    5;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 KKRIMHC 7
           ||::|||
Db       6 KKKVMHC 12


RESULT   7
Q9LML0
ID   Q9LML0      PRELIMINARY;      PRT;   512 AA.
AC   Q9LML0;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   F10K1.10 PROTEIN.
GN   F10K1.10.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RA   Liu S.X., Chan A., Yu G., Etgu P., Lee J.M., Lenz C., Pham P.,
RA   Sakano H., Toriumi M., Chung M., Goldsmith A., Liu A., Smith A.,
RA   Vaysberg M., Altafi H., Brooks S., Buehler E., Chao Q., Conn L.,
RA   Conway A., Hansen N., Johnson-Hopson C., Khan S., Kim C., Lam B.,
RA   Miranda M., Nguyen M., Palm C.J., Shinn P., Southwick A., Davis R.W.,
RA   Ecker J.R., Federspiel N.A., Theologis A.;
RT   "The sequence of BAC F10K1 from Arabidopsis thaliana chromosome 1.";
RL   Submitted (APR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RA   Theologis A.;
RL   Submitted (APR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RA   Theologis A.;
RL   Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RA   Theologis;
RL   Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- COFACTOR: FAD (BY SIMILARITY).
CC   -!- SIMILARITY: TO PYRIDINE NUCLEOTIDE-DISULPHIDE OXIDOREDUCTASES
CC       CLASS-I.
DR   EMBL; AC067971; AAF82202.1; -.
DR   InterPro; IPR001327; FAD_pyr_redox.
DR   Pfam; PF00070; pyr_redox; 1.
KW   FAD; Flavoprotein; Oxidoreductase; Redox-active center.
SQ   SEQUENCE   512 AA;  56857 MW;  1F63AFA9A1A2C13B CRC64;


  Query Match             87.8%;  Score 36;  DB 10;  Length 512;
  Best Local Similarity   71.4%;  Pred. No. 7.5;
  Matches    5;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 KKRIMHC 7
           |||::||
Db     236 KKRLLHC 242


RESULT   8
Q9D6S2
ID   Q9D6S2      PRELIMINARY;      PRT;   508 AA.
AC   Q9D6S2;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   2'-5' OLIGOADENYLATE SYNTHETASE-LIKE.
GN   OASL.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=TONGUE;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
```

```
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
DR   EMBL; AK010034; BAB26655.1; -.
DR   MGD; MGI:1344390; Oasl.
DR   InterPro; IPR001797; 25A_synth.
DR   InterPro; IPR001201; PAP_25A_core.
DR   InterPro; IPR000626; Ubiquitin.
DR   Pfam; PF00240; ubiquitin; 1.
DR   SMART; SM00213; UBQ; 1.
DR   PROSITE; PS00833; 25A_SYNTH_2; UNKNOWN_1.
DR   PROSITE; PS50152; 25A_SYNTH_3; 1.
DR   PROSITE; PS50053; UBIQUITIN_2; 1.
SQ   SEQUENCE   508 AA;  58767 MW;  9552B4540CC801A0 CRC64;


  Query Match             85.4%;  Score 35;  DB 11;  Length 508;
  Best Local Similarity   71.4%;  Pred. No. 12;
  Matches    5;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 KKRIMHC 7
           |||::||
Db     109 KKRLIHC 115


RESULT   9
Q9QXV3
ID   Q9QXV3      PRELIMINARY;      PRT;   279 AA.
AC   Q9QXV3; Q9QXV4; Q9QUP8; Q9QZX3;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   ING1 PROTEIN.
GN   ING1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC   STRAIN=129/SVJ; TISSUE=BRAIN, EMBRYONIC FIBROBLAST, AND SPLEEN;
RX   MEDLINE=20011419; PubMed=10542254;
RA   Zeremski M., Hill J.E., Kwek S.S.S., Grigorian I.A., Gurova K.V.,
RA   Garkavtsev I.V., Diatchenko L., Koonin E.V., Gudkov A.V.;
RT   "Structure and regulation of the mouse ing1 gene. Three alternative
RT   transcripts encode two PHD finger proteins that have opposite effects
RT   on p53 function.";
RL   J. Biol. Chem. 274:32172-32181(1999).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RA   Rancourt D., Garkavtsev I.;
RT   "Structural organization and expression pattern of the murine ING1
RT   gene.";
RL   Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: ISOFORM 1 INHIBITS P53-DEPENDENT TRANSCRIPTIONAL
CC       ACTIVATION AND MAY FUNCTION AS AN ONCOPROTEIN. ISOFORM 2 ACTS AS A
CC       NEGATIVE GROWTH REGULATOR BY COOPERATING WITH P53 IN
CC       TRANSCRIPTIONAL ACTIVATION OF P53-RESPONSIVE GENES AND MAY ACT AS
CC       A TUMOR SUPPRESSOR.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1 (SHOWN HERE) AND 2; ARE
CC       PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: IN THE ADULT, WIDELY EXPRESSED WITH HIGHEST
CC       LEVELS IN THYMUS AND TESTIS. EXPRESSED THROUGHOUT THE WHOLE EMBRYO
CC       AT ALL STAGES OF DEVELOPMENT EXAMINED. AT DAY 10, HIGHEST
CC       EXPRESSION IS FOUND IN THE YOLK SAC WHILE AT DAY 16 AND 18, HIGHER
CC       LEVELS ARE FOUND IN INNER COMPARTMENTS OF BONE.
CC   -!- DEVELOPMENTAL STAGE: IN THE EMBRYO, HIGHEST EXPRESSION OF ISOFORM
CC       1 IS FOUND AT DAY 11 WHILE HIGHEST EXPRESSION OF ISOFORM 2 IS
CC       FOUND AT DAY 7.
CC   -!- SIMILARITY: CONTAINS 1 PHD-FINGER DOMAIN.
DR   EMBL; AF177753; AAF16911.1; -.
DR   EMBL; AF177755; AAF16908.1; -.
DR   EMBL; AF177756; AAF16909.1; -.
DR   EMBL; AF177757; AAF16910.1; -.
DR   EMBL; AF149820; AAF09183.1; -.
DR   MGD; MGI:1349481; Ing1.
DR   InterPro; IPR001965; PHD.
DR   Pfam; PF00628; PHD; 1.
DR   SMART; SM00249; PHD; 1.
KW   Oncogene; Anti-oncogene; Alternative splicing.
FT   DOMAIN      210    259       PHD-FINGER.
FT   VARSPLIC      1     94       MISSING (IN ISOFORM 2).
FT   CONFLICT    203    203       L -> F (IN REF. 2).
SQ   SEQUENCE   279 AA;  32109 MW;  6765C984EEF179F4 CRC64;


  Query Match             82.9%;  Score 34;  DB 11;  Length 279;
  Best Local Similarity   57.1%;  Pred. No. 12;
  Matches    4;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 KKRIMHC 7
           |:|::||
Db      67 KRRVLHC 73


RESULT  10
Q9M149
ID   Q9M149      PRELIMINARY;      PRT;   401 AA.
AC   Q9M149;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   PUTATIVE PHOSPHATIDYLINOSITOL KINASE.
GN   AT4G01190.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Lamar B., Stoneking T., Stumpf J., Mewes H.W., Lemcke K.,
RA   Mayer K.F.X.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   EU Arabidopsis sequencing project;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL161491; CAB80928.1; -.
DR   InterPro; IPR002498; PIP5K.
DR   Pfam; PF01504; PIP5K; 1.
DR   SMART; SM00330; PIPKc; 1.
KW   Kinase.
SQ   SEQUENCE   401 AA;  45658 MW;  8A12D10DA2DED4CA CRC64;


  Query Match             82.9%;  Score 34;  DB 10;  Length 401;
  Best Local Similarity   85.7%;  Pred. No. 16;
  Matches    6;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 KKRIMHC 7
           |||| ||
Db     353 KKRIEHC 359
```

```
RESULT  11
O04613
ID   O04613      PRELIMINARY;      PRT;   431 AA.
AC   O04613;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   SIMILAR TO PHOSPHATIDYLINOSITOL-4-PHOSPHATE -KINASE TYPE II.
GN   A_IG002N01.9.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RA   Scheet P., Maggi L.;
RL   Submitted (JUN-1997) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RA   Waterston R.;
RL   Submitted (JUN-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF007269; AAB61030.1; -.
DR   Mendel; 17574; Arath;2860;17574.
DR   InterPro; IPR002498; PIP5K.
DR   Pfam; PF01504; PIP5K; 1.
DR   SMART; SM00330; PIPKc; 1.
SQ   SEQUENCE   431 AA;  49356 MW;  07A53F23BDDD942B CRC64;

  Query Match             82.9%;  Score 34;  DB 10;  Length 431;
  Best Local Similarity   85.7%;  Pred. No. 17;
  Matches    6;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 KKRIMHC 7
           |||| ||
Db     383 KKRIEHC 389


RESULT  12
Q03949
ID   Q03949      PRELIMINARY;      PRT;   189 AA.
AC   Q03949;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   INSERTION ELEMENT IS895 HYPOTHETICAL 21.9 KDA PROTEIN (ORFI).
OS   Anabaena sp. (strain PCC 7120).
OC   Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX   NCBI_TaxID=103690;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91358370; PubMed=1653219;
RA   Alam J., Vrba J.M., Cai Y., Martin J.A., Weislo L.J., Curtis S.E.;
RT   "Characterization of the IS895 family of insertion sequences from the
RT   cyanobacterium Anabaena sp. strain PCC 7120.";
RL   J. Bacteriol. 173:5778-5783(1991).
DR   EMBL; M67475; AAA98138.1; -.
DR   PIR; A38117; A38117.
KW   Hypothetical protein; Transposable element.
SQ   SEQUENCE   189 AA;  21937 MW;  B873A342856C2103 CRC64;

  Query Match             80.5%;  Score 33;  DB 2;  Length 189;
  Best Local Similarity   71.4%;  Pred. No. 14;
  Matches    5;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 KKRIMHC 7
           |||: ||
Db     171 KKRLKHC 177


RESULT  13
Q9AKS0
ID   Q9AKS0      PRELIMINARY;      PRT;   267 AA.
AC   Q9AKS0;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   TATC PROTEIN.
GN   TATC.
OS   Pseudomonas stutzeri (Pseudomonas perfectomarina).
OC   Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC   Pseudomonas.
OX   NCBI_TaxID=316;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ZOBELL ATCC14405;
RX   MEDLINE=21101859; PubMed=11160097;
RA   Heikkilae M.P., Honisch U., Wunsch P., Zumft W.G.;
RT   "Role of the Tat transport system in nitrous oxide reductase
RT   translocation and cytochrome cd1 biosyntheis in Pseudomonas
RT   stutzeri.";
RL   J. Bacteriol. 183:1663-1671(2001).
DR   EMBL; AJ299712; CAC29149.1; -.
SQ   SEQUENCE   267 AA;  29439 MW;  4369EB5E38BCD736 CRC64;

  Query Match             80.5%;  Score 33;  DB 2;  Length 267;
  Best Local Similarity   57.1%;  Pred. No. 19;
  Matches    4;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 KKRIMHC 7
           :||::||
Db      21 RKRLLHC 27


RESULT  14
Q9ST63
ID   Q9ST63      PRELIMINARY;      PRT;   495 AA.
AC   Q9ST63;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   PUTATIVE INTERNAL ROTENONE-INSENSITIVE NADH DEHYDROGENASE.
GN   NDA1.
OS   Solanum tuberosum (Potato).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC   Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
OX   NCBI_TaxID=4113;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. DESIREE; TISSUE=LEAF;
RA   Rasmusson A.G., Svensson A., Knoop V., Grohmann L., Brennicke A.;
RT   "Homologues of yeast and bacterial rotenone-insensitive NADH
RT   dehydrogenases in higher eukaryotes: two enzymes are present in potato
RT   mitochondria.";
RL   Submitted (AUG-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- COFACTOR: FAD (BY SIMILARITY).
CC   -!- SIMILARITY: TO PYRIDINE NUCLEOTIDE-DISULPHIDE OXIDOREDUCTASES
CC       CLASS-I.
DR   EMBL; AJ245861; CAB52796.1; -.
DR   InterPro; IPR001327; FAD_pyr_redox.
DR   Pfam; PF00070; pyr_redox; 1.
KW   FAD; Flavoprotein; NAD; Oxidoreductase; Redox-active center.
SQ   SEQUENCE   495 AA;  54902 MW;  6AFFC807BEB01340 CRC64;

  Query Match             80.5%;  Score 33;  DB 10;  Length 495;
  Best Local Similarity   57.1%;  Pred. No. 32;
  Matches    4;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy       1 KKRIMHC 7
           |:|::||
Db     224 KRRLLHC 230


RESULT  15
O80874
ID   O80874      PRELIMINARY;      PRT;   508 AA.
AC   O80874;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   PUTATIVE UBIQINONE REDUCTASE.
GN   F23F1.9.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RA   Rounsley S.D., Lin X., Ketchum K.A., Crosby M.L., Brandon R.C.,
RA   Sykes S.M., Kaul S., Mason T.M., Kerlavage A.R., Adams M.D.,
RA   Somerville C.R., Venter J.C.;
RT   "Arabidopsis thaliana chromosome II BAC F23F1 genomic sequence.";
RL   Submitted (AUG-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- COFACTOR: FAD (BY SIMILARITY).
CC   -!- SIMILARITY: TO PYRIDINE NUCLEOTIDE-DISULPHIDE OXIDOREDUCTASES
CC       CLASS-I.
DR   EMBL; AC004680; AAC31853.1; -.
DR   InterPro; IPR001327; FAD_pyr_redox.
DR   Pfam; PF00070; pyr_redox; 1.
KW   FAD; Flavoprotein; Oxidoreductase; Redox-active center.
SQ   SEQUENCE   508 AA;  56503 MW;  266A434E702A0C27 CRC64;


  Query Match             80.5%;  Score 33;  DB 10;  Length 508;
  Best Local Similarity   57.1%;  Pred. No. 33;
  Matches    4;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 KKRIMHC 7
           |:|::||
Db     236 KRRLLHC 242


Search completed: February 12, 2002, 12:03:25
Job time: 816 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: February 12, 2002, 12:04:01 ; Search time 30.28 Seconds
(without alignments)
8.476 Million cell updates/sec

Title: US-09-606-129A-18
Perfect score: 41
Sequence: 1 KKRIMHC 7

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	41	100.0	295	1	BIEA_RAT	P46844 rattus norv
2	38	92.7	296	1	BIEA_HUMAN	P53004 homo sapien
3	35	85.4	473	1	OASL_MOUSE	Q9z2f2 mus musculu
4	33	80.5	925	1	PIP1_YEAST	P40020 saccharomyc
5	32	78.0	140	1	FKB2_MOUSE	P45878 mus musculu
6	32	78.0	141	1	FKB2_HUMAN	P26885 homo sapien
7	32	78.0	328	1	HYBA_ECOLI	P37179 escherichia
8	32	78.0	528	1	DYR2_HUMAN	Q92630 homo sapien
9	32	78.0	571	1	AMP1_LYCES	Q10712 lycopersico
10	32	78.0	573	1	AMPL_SOLTU	P31427 solanum tub
11	31	75.6	105	1	RL44_TRYBB	P17843 trypanosoma
12	31	75.6	279	1	Y091_NPVOP	O10341 orgyia pseu
13	31	75.6	323	1	CYCH_XENLA	P51947 xenopus lae
14	31	75.6	451	1	SUN_HAEIN	P44788 haemophilus
15	31	75.6	768	1	CUL3_HUMAN	Q13618 homo sapien
16	31	75.6	768	1	CUL3_MOUSE	Q9jlv5 mus musculu
17	31	75.6	861	1	UL52_HSV7J	P52468 human herpe
18	30	73.2	108	1	YZ09_METJA	Q60271 methanococc
19	30	73.2	4128	1	PRKD_MOUSE	P97313 mus musculu
20	29	70.7	38	1	RL36_THEMA	Q9xli6 thermotoga
21	29	70.7	142	1	SECB_BUCAI	P57161 buchnera ap
22	29	70.7	144	1	YHP5_YEAST	P38808 saccharomyc
23	29	70.7	257	1	YZG1_CAEEL	P55326 caenorhabdi
24	29	70.7	300	1	RANT_BPP22	P03037 bacteriopha
25	29	70.7	359	1	ODPB_RAT	P49432 rattus norv
26	29	70.7	423	1	SHP1_YEAST	P34223 saccharomyc
27	29	70.7	560	1	YD2H_SCHPO	Q10264 schizosacch
28	29	70.7	588	1	DYR3_HUMAN	O43781 homo sapien
29	29	70.7	632	1	AFUB_HAEIN	Q57341 haemophilus
30	29	70.7	670	1	REP_HAEIN	P44804 haemophilus
31	29	70.7	918	1	YK62_CAEEL	P34341 caenorhabdi
32	29	70.7	1059	1	CAPU_DROME	Q24120 drosophila
33	29	70.7	1790	1	SEPA_EMENI	P78621 emericella
34	28	68.3	113	1	FLIT_BACSU	P39740 bacillus su
35	28	68.3	172	1	IPYR_RICPR	Q9zcw5 rickettsia
36	28	68.3	188	1	VEGB_MOUSE	P49766 mus musculu
37	28	68.3	249	1	YAT6_RHOBL	P05449 rhodopseudo
38	28	68.3	313	1	YQAM_BACSU	P45910 bacillus su
39	28	68.3	324	1	IDSA_METTM	Q53479 methanobact
40	28	68.3	335	1	UL16_HSV6U	P24442 human herpe
41	28	68.3	356	1	GBA1_DICDI	P16894 dictyosteli
42	28	68.3	428	1	DGTP_MYCSM	O52199 mycobacteri
43	28	68.3	459	1	NEK3_HUMAN	P51956 homo sapien
44	28	68.3	477	1	URT1_DESRO	P98119 desmodus ro
45	28	68.3	477	1	URT2_DESRO	P15638 desmodus ro

ALIGNMENTS

RESULT 1
BIEA_RAT
ID BIEA_RAT STANDARD; PRT; 295 AA.
AC P46844;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE BILIVERDIN REDUCTASE A PRECURSOR (EC 1.3.1.24) (BILIVERDIN-IX ALPHA-
DE REDUCTASE).
GN BLVRA OR BLVR.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Kidney;
RX MEDLINE=92156147; PubMed=1371282;
RA Fakhrai H., Maines M.D.;
RT "Expression and characterization of a cDNA for rat kidney biliverdin
RT reductase. Evidence suggesting the liver and kidney enzymes are the
RT same transcript product.";
RL J. Biol. Chem. 267:4023-4029(1992).
RN [2]
RP MUTAGENESIS.
RX MEDLINE=94291657; PubMed=8020496;
RA McCoubrey W.K. Jr., Maines M.D.;
RT "Site-directed mutagenesis of cysteine residues in biliverdin
RT reductase. Roles in substrate and cofactor binding.";
RL Eur. J. Biochem. 222:597-603(1994).
CC -!- FUNCTION: CONVERTS BILIVERDIN TO BILIRUBIN; DISPLAYS TWO DISTINCT
CC PH OPTIMA USING A DIFFERENT COFACTOR AT EACH PH: NADH AT THE LOWER
CC PH 6.7-6.9 RANGE AND NADPH AT PH 8.5-8.7. NADPH, HOWEVER, IS THE
CC PROBABLE COFACTOR IN BIOLOGICAL SYSTEMS.
CC -!- CATALYTIC ACTIVITY: BILIRUBIN + NAD(P)(+) = BILIVERDIN + NAD(P)H.
CC -!- COFACTOR: BINDS ONE ZINC ION.
CC -!- PATHWAY: FINAL STEP IN HEME METABOLISM.
CC -!- SUBUNIT: MONOMER (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -!- SIMILARITY: TO E.COLI YHHX.
CC ---

CC ---
DR EMBL; M81681; AAA40830.1; -.
DR InterPro; IPR000683; GFO_IDH_MocA.
DR Pfam; PF01408; GFO_IDH_MocA; 1.
KW Oxidoreductase; NAD; NADP; Zinc.
FT PROPEP 1 2
FT CHAIN 3 295 BILIVERDIN REDUCTASE A.
FT DOMAIN 11 16 POLY-VAL.

```
FT   METAL       279    279       ZINC (POTENTIAL).
FT   METAL       280    280       ZINC (POTENTIAL).
FT   METAL       291    291       ZINC (POTENTIAL).
FT   METAL       292    292       ZINC (POTENTIAL).
FT   MUTAGEN      73     73       C->A: LOSS OF ACTIVITY.
FT   MUTAGEN     280    280       C->A: REDUCED ACTIVITY.
FT   MUTAGEN     291    291       C->A: REDUCED ACTIVITY.
SQ   SEQUENCE   295 AA;  33565 MW;  219C8EA96C150588 CRC64;


  Query Match             100.0%;  Score 41;  DB 1;  Length 295;
  Best Local Similarity   100.0%;  Pred. No. 0.13;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 KKRIMHC 7
           |||||||
Db     274 KKRIMHC 280


RESULT   2
BIEA_HUMAN
ID   BIEA_HUMAN     STANDARD;      PRT;   296 AA.
AC   P53004;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   BILIVERDIN REDUCTASE A PRECURSOR (EC 1.3.1.24) (BILIVERDIN-IX ALPHA-
DE   REDUCTASE).
GN   BLVRA OR BLVR OR BVR.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Placenta;
RX   MEDLINE=96202961; PubMed=8631357;
RA   Maines M.D., Polevoda B.V., Huang T.-J., McCoubrey W.K. Jr.;
RT   "Human biliverdin IXalpha reductase is a zinc-metalloprotein.
RT   Characterization of purified and Escherichia coli expressed
RT   enzymes.";
RL   Eur. J. Biochem. 235:372-381(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Komuro A., Tobe T., Nakano Y., Yamaguchi T., Tomita M.;
RL   Submitted (JAN-1996) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 1-117 FROM N.A.
RA   Cordes M., Wollam C., Carter T.;
RL   Submitted (JUN-1998) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE OF 3-36; 48-74 AND 228-248.
RC   TISSUE=Liver;
RX   MEDLINE=93143333; PubMed=8424666;
RA   Maines M.D., Trakshel G.M.;
RT   "Purification and characterization of human biliverdin reductase.";
RL   Arch. Biochem. Biophys. 300:320-326(1993).
RN   [5]
RP   SEQUENCE OF 3-22.
RC   TISSUE=Liver;
RX   MEDLINE=95014177; PubMed=7929092;
RA   Yamaguchi T., Komoda Y., Nakajima H.;
RT   "Biliverdin-IX alpha reductase and biliverdin-IX beta reductase from
RT   human liver. Purification and characterization.";
RL   J. Biol. Chem. 269:24343-24348(1994).
CC   -!- FUNCTION: CONVERTS BILIVERDIN TO BILIRUBIN.
CC   -!- CATALYTIC ACTIVITY: BILIRUBIN + NAD(P)(+) = BILIVERDIN + NAD(P)H.
CC   -!- COFACTOR: BINDS ONE ZINC ION. HAS DUAL PH/COFACTOR (NADH, NADPH)
CC       SPECIFICITY. USES NADH AT THE ACIDIC PH RANGE (6-6.7) AND NADPH AT
CC       THE ALKALINE RANGE (8.5-8.7).
CC   -!- PATHWAY: FINAL STEP IN HEME METABOLISM.
CC   -!- SUBUNIT: MONOMER.
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC   -!- TISSUE SPECIFICITY: LIVER.
CC   -!- SIMILARITY: TO E.COLI YHHX.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X93086; CAA63635.1; -.
DR   EMBL; U34877; AAC35588.1; -.
DR   EMBL; AC005189; AAC25526.1; -.
DR   MIM; 109750; -.
DR   InterPro; IPR000683; GFO_IDH_MocA.
DR   Pfam; PF01408; GFO_IDH_MocA; 1.
KW   Oxidoreductase; NAD; NADP; Zinc.
FT   PROPEP        1      2
FT   CHAIN         3    296       BILIVERDIN REDUCTASE A.
FT   DOMAIN       11     16       POLY-VAL.
FT   METAL       280    280       ZINC (POTENTIAL).
FT   METAL       281    281       ZINC (POTENTIAL).
FT   METAL       292    292       ZINC (POTENTIAL).
FT   METAL       293    293       ZINC (POTENTIAL).
FT   CONFLICT      3      3       A -> T (IN REF. 2).
FT   CONFLICT    154    155       SD -> AG (IN REF. 2).
FT   CONFLICT    160    160       D -> E (IN REF. 2).
SQ   SEQUENCE   296 AA;  33488 MW;  0DFD3B386F4DFC0A CRC64;


  Query Match             92.7%;  Score 38;  DB 1;  Length 296;
  Best Local Similarity   85.7%;  Pred. No. 0.55;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 KKRIMHC 7
           ||||:||
Db     275 KKRILHC 281


RESULT   3
OASL_MOUSE
ID   OASL_MOUSE     STANDARD;      PRT;   473 AA.
AC   Q9Z2F2;
DT   20-AUG-2001 (Rel. 40, Created)
DT   20-AUG-2001 (Rel. 40, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   54 KDA 2'-5'-OLIGOADENYLATE SYNTHETASE LIKE PROTEIN (EC 2.7.7.-) (P54
DE   OASL) (P54OASL) (M1204).
GN   OASL.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Spleen;
RX   MEDLINE=99323969; PubMed=10395668;
RA   Tiefenthaler M., Marksteiner R., Neyer S., Koch F., Hofer S.,
RA   Schuler G., Nussenzweig M., Schneider R., Heufler C.;
RT   "M1204, a novel 2',5' oligoadenylate synthetase with a ubiquitin-like
RT   extension, is induced during maturation of murine dendritic cells.";
RL   J. Immunol. 163:760-765(1999).
CC   -!- FUNCTION: MAY PLAY A ROLE IN MEDIATING RESISTANCE TO VIRUS
CC       INFECTION, CONTROL OF CELL GROWTH, DIFFERENTIATION, AND APOPTOSIS.
CC   -!- CATALYTIC ACTIVITY: BINDS DOUBLE-STRANDED RNA AND POLYMERIZES ATP
CC       INTO PPP(A2'P5'A)N OLIGOMERS, WHICH ACTIVATE THE LATENT RNASE L
CC       THAT, WHEN ACTIVATED, CLEAVES SINGLE-STRANDED RNAS.
CC   -!- TISSUE SPECIFICITY: STRONGLY EXPRESSED IN SPLEEN DENDRITIC CELLS,
CC       WHEREAS, IN BONE MARROW-DERIVED DENDRITIC CELLS, THE AMOUNT
CC       INCREASES DURING THE MATURATION PROCESS. EXPRESSED IN MANY ORGANS,
CC       THE HIGHEST LEVELS BEING IN THYMUS, LUNG, AND BONE MARROW.
CC   -!- SIMILARITY: BELONGS TO THE 2-5A SYNTHETASE FAMILY.
```

```
CC   -!- SIMILARITY: CONTAINS 1 UBIQUITIN-LIKE DOMAIN.
CC   -!- CAUTION: THIS MAY NOT BE THE TRUE ORTHOLOG OF HUMAN OASL.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF068835; AAD02818.1; -.
DR   MGD; MGI:1344390; Oasl.
DR   InterPro; IPR001797; 25A_synth.
DR   InterPro; IPR001201; PAP_25A_core.
DR   InterPro; IPR000626; Ubiquitin.
DR   Pfam; PF00240; ubiquitin; 1.
DR   PROSITE; PS00832; 25A_SYNTH_1; FALSE_NEG.
DR   PROSITE; PS00833; 25A_SYNTH_2; 1.
DR   PROSITE; PS50152; 25A_SYNTH_3; 1.
DR   PROSITE; PS50053; UBIQUITIN_2; FALSE_NEG.
KW   RNA-binding; Transferase; Nucleotidyltransferase.
FT   DOMAIN      435    473       UBIQUITIN-LIKE.
SQ   SEQUENCE   473 AA;  54625 MW;  570E0E08A51C8460 CRC64;


  Query Match             85.4%;  Score 35;  DB 1;  Length 473;
  Best Local Similarity   71.4%;  Pred. No. 3.8;
  Matches    5;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 KKRIMHC 7
          |||::||
Db    109 KKRLIHC 115


RESULT   4
PIP1_YEAST
ID   PIP1_YEAST     STANDARD;      PRT;   925 AA.
AC   P40020;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   POLYMERASE-INTERACTING PROTEIN 1 (FACTOR INTERACTING WITH REF).
GN   PIP1 OR FIR1 OR YER032W.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W303;
RA   del Olmo M., Gross S., Moore C.L.;
RL   Submitted (FEB-1997) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / AB972;
RA   Dietrich F.S., Mulligan J.T., Hennessey K.M., Allen E., Araujo R.,
RA   Aviles E., Berno A., Brennan T., Carpenter J., Chen E., Cherry J.M.,
RA   Chung E., Duncan M., Guzman E., Hartzell G., Hunicke-Smith S.,
RA   Hyman R., Kayser A., Komp C., Lashkari D., Lew H., Lin D.,
RA   Mosedale D., Nakahara K., Namath A., Norgren R., Oefner P., Oh C.,
RA   Petel F.X., Roberts D., Sehl P., Schramm S., Shogren T., Smith V.,
RA   Taylor P., Wei Y., Yelton M., Botstein D., Davis R.W.;
RL   Submitted (DEC-1994) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   CHARACTERIZATION.
RX   MEDLINE=97339480; PubMed=9196079;
RA   Russnak R., Pereira S., Platt T.;
RT   "RNA binding analysis of yeast REF2 and its two-hybrid interaction
RT   with a new gene product, FIR1.";
RL   Gene Expr. 6:241-258(1996).
CC   -!- FUNCTION: INTERACTS WITH POLY(A) POLYMERASE AND WITH REF2.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U17262; AAB46625.1; -.
DR   EMBL; U18778; AAB64565.1; -.
DR   SGD; S0000834; FIR1.
FT   CONFLICT    663    663       R -> P (IN REF. 2).
SQ   SEQUENCE   925 AA;  104701 MW;  707D9839EE31322B CRC64;


  Query Match             80.5%;  Score 33;  DB 1;  Length 925;
  Best Local Similarity   71.4%;  Pred. No. 19;
  Matches    5;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 KKRIMHC 7
          |||: ||
Db    841 KKRLSHC 847


RESULT   5
FKB2_MOUSE
ID   FKB2_MOUSE     STANDARD;      PRT;   140 AA.
AC   P45878;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   FK506-BINDING PROTEIN PRECURSOR (FKBP-13) (PEPTIDYL-PROLYL CIS-TRANS
DE   ISOMERASE) (PPIASE) (EC 5.2.1.8).
GN   FKBP2 OR FKBP13.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=129/SVJ; TISSUE=Liver;
RX   MEDLINE=94085790; PubMed=7505249;
RA   Hendrickson B.A., Zhang W., Craig R.J., Jin Y.J., Bierer R.E.,
RA   Burakoff S.J., Dilella A.G.;
RT   "Structural organization of the genes encoding human and murine
RT   FK506-binding protein (FKBP) 13 and comparison to FKBP1.";
RL   Gene 134:271-275(1993).
CC   -!- FUNCTION: PPIASES ACCELERATE THE FOLDING OF PROTEINS.
CC   -!- CATALYTIC ACTIVITY: CIS-TRANS ISOMERIZATION OF PROLINE IMIDIC
CC       PEPTIDE BONDS IN OLIGOPEPTIDES.
CC   -!- ENZYME REGULATION: INHIBITED BY BOTH FK506 AND RAPAMYCIN.
CC   -!- SUBCELLULAR LOCATION: ENDOPLASMIC RETICULUM LUMEN. MEMBRANE
CC       ASSOCIATED (PROBABLE).
CC   -!- SIMILARITY: BELONGS TO THE FKBP-TYPE PPIASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M77831; AAA37631.1; -.
DR   HSSP; P20081; 1YAT.
DR   MGD; MGI:95542; Fkbp2.
DR   InterPro; IPR001179; FKBP_PPIase.
DR   Pfam; PF00254; FKBP; 1.
DR   PROSITE; PS00453; FKBP_PPIASE_1; 1.
DR   PROSITE; PS00454; FKBP_PPIASE_2; 1.
DR   PROSITE; PS50059; FKBP_PPIASE_3; 1.
KW   Isomerase; Rotamase; Signal; Endoplasmic reticulum.
FT   SIGNAL      1     22       POTENTIAL.
```

```
FT   CHAIN        23    140       FK506-BINDING PROTEIN.
FT   SITE        137    140       PREVENT SECRETION FROM ER (POTENTIAL).
SQ   SEQUENCE   140 AA;  15344 MW;  F4E7FCC7766A0416 CRC64;


  Query Match             78.0%;  Score 32;  DB 1;  Length 140;
  Best Local Similarity   71.4%;  Pred. No. 4.8;
  Matches    5;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 KKRIMHC 7
             |||: ||
Db      34 KKRVDHC 40


RESULT   6
FKB2_HUMAN
ID   FKB2_HUMAN     STANDARD;      PRT;   141 AA.
AC   P26885;
DT   01-AUG-1992 (Rel. 23, Created)
DT   01-AUG-1992 (Rel. 23, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   FK506-BINDING PROTEIN PRECURSOR (FKBP-13) (PEPTIDYL-PROLYL CIS-TRANS
DE   ISOMERASE) (PPIASE) (EC 5.2.1.8).
GN   FKBP2 OR FKBP13.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Colon carcinoma;
RX   MEDLINE=91319747; PubMed=1713687;
RA   Jin Y.-J., Albers M.W., Lane W.S., Bierer B.E., Schreiber S.L.,
RA   Burakoff S.J.;
RT   "Molecular cloning of a membrane-associated human FK506- and
RT   rapamycin-binding protein, FKBP-13.";
RL   Proc. Natl. Acad. Sci. U.S.A. 88:6677-6681(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93112052; PubMed=1281998;
RA   Dilella A.G., Hawkins A., Craig R.J., Schreiber S.L., Griffin C.A.;
RT   "Chromosomal band assignments of the genes encoding human FKBP12 and
RT   FKBP13.";
RL   Biochem. Biophys. Res. Commun. 189:819-823(1992).
CC   -!- FUNCTION: PPIASES ACCELERATE THE FOLDING OF PROTEINS.
CC   -!- CATALYTIC ACTIVITY: CIS-TRANS ISOMERIZATION OF PROLINE IMIDIC
CC       PEPTIDE BONDS IN OLIGOPEPTIDES.
CC   -!- ENZYME REGULATION: INHIBITED BY BOTH FK506 AND RAPAMYCIN.
CC   -!- SUBCELLULAR LOCATION: ENDOPLASMIC RETICULUM LUMEN. MEMBRANE
CC       ASSOCIATED (PROBABLE).
CC   -!- TISSUE SPECIFICITY: T-CELLS AND THYMUS.
CC   -!- SIMILARITY: BELONGS TO THE FKBP-TYPE PPIASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M65128; AAA58473.1; -.
DR   EMBL; M75099; AAA36563.1; -.
DR   PIR; JC1365; JC1365.
DR   HSSP; P20071; 1FKT.
DR   MIM; 186946; -.
DR   InterPro; IPR001179; FKBP_PPIase.
DR   Pfam; PF00254; FKBP; 1.
DR   PROSITE; PS00453; FKBP_PPIASE_1; 1.
DR   PROSITE; PS00454; FKBP_PPIASE_2; 1.
DR   PROSITE; PS50059; FKBP_PPIASE_3; 1.
KW   Isomerase; Rotamase; Signal; Endoplasmic reticulum; Polymorphism.
FT   SIGNAL        1     21
FT   CHAIN        22    141       FK506-BINDING PROTEIN.
FT   SITE        138    141       PREVENT SECRETION FROM ER (POTENTIAL).
FT   VARIANT      21     21       S -> TA.
FT                                /FTId=VAR_006410.
FT   VARIANT      24     24       T -> A.
FT                                /FTId=VAR_006411.
FT   VARIANT      96     96       Y -> C.
FT                                /FTId=VAR_006412.
SQ   SEQUENCE   141 AA;  15654 MW;  9F4751CA7D82D064 CRC64;


  Query Match             78.0%;  Score 32;  DB 1;  Length 141;
  Best Local Similarity   71.4%;  Pred. No. 4.9;
  Matches    5;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 KKRIMHC 7
             |||: ||
Db      35 KKRVDHC 41


RESULT   7
HYBA_ECOLI
ID   HYBA_ECOLI     STANDARD;      PRT;   328 AA.
AC   P37179;
DT   01-OCT-1994 (Rel. 30, Created)
DT   01-OCT-1994 (Rel. 30, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   HYDROGENASE-2 OPERON PROTEIN HYBA PRECURSOR.
GN   HYBA OR B2996 OR Z4350 OR ECS3881.
OS   Escherichia coli, and
OS   Escherichia coli O157:H7.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562, 83334;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / TG1;
RX   MEDLINE=94292472; PubMed=8021226;
RA   Menon N.K., Chatelus C.Y., Dervartanian M., Wendt J.C.,
RA   Shanmugam K.T., Peck H.D. Jr., Przybyla A.E.;
RT   "Cloning, sequencing, and mutational analysis of the hyb operon
RT   encoding Escherichia coli hydrogenase 2.";
RL   J. Bacteriol. 176:4416-4423(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=97426617; PubMed=9278503;
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX   MEDLINE=21074935; PubMed=11206551;
RA   Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA   Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA   Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA   Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA   Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA   Welch R.A., Blattner F.R.;
RT   "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL   Nature 409:529-533(2001).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O157:H7 / RIMD 0509952;
RX   MEDLINE=21156231; PubMed=11258796;
RA   Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA   Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA   Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA   Kuhara S., Shiba T., Hattori M., Shinagawa H.;
```

```
RT   "Complete genome sequence of enterohemorrhagic Escherichia coli
RT   O157:H7 and genomic comparison with a laboratory strain K-12.";
RL   DNA Res. 8:11-22(2001).
CC   -!- FUNCTION: PARTICIPATES IN THE PERIPLASMIC ELECTRON-TRANSFERRING
CC       ACTIVITY OF HYDROGENASE 2 DURING ITS CATALYTIC TURNOVER.
CC   -!- COFACTOR: BINDS 3 4FE-4S CLUSTERS AND A 3FE-4S CLUSTER.
CC   -!- SUBCELLULAR LOCATION: PERIPLASMIC.
CC   -!- SIMILARITY: THE IRON-SULFUR CENTERS ARE SIMILAR TO THOSE OF
CC       'BACTERIAL-TYPE' 4FE-4S FERREDOXINS.
CC   -!- CAUTION: WAS ORIGINALLY (REF.1) THOUGHT TO BE THE SMALL SUBUNIT
CC       OF HYDROGENASE 2.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U09177; AAA21589.1; -.
DR   EMBL; U28377; AAA69163.1; -.
DR   EMBL; AE000382; AAC76032.1; -.
DR   EMBL; AE005529; AAG58133.1; -.
DR   EMBL; AP002563; BAB37304.1; -.
DR   HSSP; P55907; 1XER.
DR   EcoGene; EG11799; hybA.
DR   InterPro; IPR001450; 4FE4S_ferrdxin.
DR   Pfam; PF00037; fer4; 1.
DR   PROSITE; PS00198; 4FE4S_FERREDOXIN; 1.
KW   Oxidoreductase; Signal; Periplasmic; Iron-sulfur; 4Fe-4S; 3Fe-4S;
KW   Complete proteome.
FT   SIGNAL       1     27       POTENTIAL.
FT   CHAIN       28    328       HYDROGENASE-2 OPERON PROTEIN HYBA.
FT   METAL       47     47       IRON-SULFUR 1 (4FE-4S) (POTENTIAL).
FT   METAL       50     50       IRON-SULFUR 1 (4FE-4S) (POTENTIAL).
FT   METAL       53     53       IRON-SULFUR 1 (4FE-4S) (POTENTIAL).
FT   METAL       57     57       IRON-SULFUR 1 (4FE-4S) (POTENTIAL).
FT   METAL      112    112       IRON-SULFUR 2 (3FE-4S) (POTENTIAL).
FT   METAL      115    115       IRON-SULFUR 2 (3FE-4S) (POTENTIAL).
FT   METAL      120    120       IRON-SULFUR 2 (3FE-4S) (POTENTIAL).
FT   METAL      124    124       IRON-SULFUR 2 (3FE-4S) (POTENTIAL).
FT   METAL      145    145       IRON-SULFUR 3 (4FE-4S) (POTENTIAL).
FT   METAL      148    148       IRON-SULFUR 3 (4FE-4S) (POTENTIAL).
FT   METAL      151    151       IRON-SULFUR 3 (4FE-4S) (POTENTIAL).
FT   METAL      155    155       IRON-SULFUR 3 (4FE-4S) (POTENTIAL).
FT   METAL      174    174       IRON-SULFUR 4 (4FE-4S) (POTENTIAL).
FT   METAL      177    177       IRON-SULFUR 4 (4FE-4S) (POTENTIAL).
FT   METAL      193    193       IRON-SULFUR 4 (4FE-4S) (POTENTIAL).
FT   METAL      197    197       IRON-SULFUR 4 (4FE-4S) (POTENTIAL).
SQ   SEQUENCE   328 AA;  36003 MW;  77203A0F50F61662 CRC64;


  Query Match             78.0%;  Score 32;  DB 1;  Length 328;
  Best Local Similarity   71.4%;  Pred. No. 11;
  Matches    5;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 KKRIMHC 7
           ||: |||
Db     109 KKQCMHC 115


RESULT   8
DYR2_HUMAN
ID   DYR2_HUMAN     STANDARD;      PRT;   528 AA.
AC   Q92630;
DT   20-AUG-2001 (Rel. 40, Created)
DT   20-AUG-2001 (Rel. 40, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   DUAL-SPECIFICITY TYROSINE-PHOSPHORYLATION REGULATED KINASE 2
DE   (EC 2.7.1.-).
GN   DYRK2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=98421512; PubMed=9748265;
RA   Becker W., Weber Y., Wetzel K., Eirmbter K., Tejedor F.J.,
RA   Joost H.-G.;
RT   "Sequence characteristics, subcellular localization, and substrate
RT   specificity of DYRK-related kinases, a novel family of dual
RT   specificity protein kinases.";
RL   J. Biol. Chem. 273:25893-25902(1998).
RN   [2]
RP   SEQUENCE OF 320-528 FROM N.A.
RC   TISSUE=Placenta;
RA   Becker W., Joost H.-G.;
RL   Submitted (NOV-1996) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: IN VITRO; CAN PHOSPHORYLATE HISTONES H3 AND H2B ON SER
CC       AND THR RESIDUES. MAY BE INVOLVED IN THE REGULATION OF CELLULAR
CC       GROWTH AND/OR DEVELOPMENT.
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC   -!- PTM: AUTHOPHOSPHORYLATED ON TYR RESIDUES.
CC   -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC       MNB/DYRK SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Y13493; CAA73885.1; -.
DR   EMBL; Y09216; CAA70418.1; -.
DR   HSSP; Q16539; 1WFC.
DR   MIM; 603496; -.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002290; Ser_thr_kin_actsite.
DR   Pfam; PF00069; pkinase; 2.
DR   SMART; SM00220; S_TKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
KW   Transferase; Serine/threonine-protein kinase; Tyrosine-protein kinase;
KW   ATP-binding; Phosphorylation.
FT   DOMAIN      149    462       PROTEIN KINASE.
FT   NP_BIND     155    163       ATP (BY SIMILARITY).
FT   BINDING     178    178       ATP (BY SIMILARITY).
FT   ACT_SITE    275    275       BY SIMILARITY.
SQ   SEQUENCE   528 AA;  59714 MW;  AF2C6822ED9522D7 CRC64;


  Query Match             78.0%;  Score 32;  DB 1;  Length 528;
  Best Local Similarity   71.4%;  Pred. No. 18;
  Matches    5;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 KKRIMHC 7
           | ||:||
Db     268 KNRIIHC 274


RESULT   9
AMP1_LYCES
ID   AMP1_LYCES     STANDARD;      PRT;   571 AA.
AC   Q10712; Q9S9A3;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   CHLOROPLAST AMINOPEPTIDASE 1 PRECURSOR (EC 3.4.11.1) (LEUCINE
DE   AMINOPEPTIDASE) (LAP) (LEUCYL AMINOPEPTIDASE) (PROLINE AMINOPEPTIDASE)
DE   (EC 3.4.11.5) (PROLYL AMINOPEPTIDASE) (DR57).
```

```
GN   LAPA1 OR LAP OR LAP2.
OS   Lycopersicon esculentum (Tomato).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC   Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
OX   NCBI_TaxID=4081;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. PETO 238R; TISSUE=Leaf;
RX   MEDLINE=96421572; PubMed=8824220;
RA   Gu Y.Q., Chao W.S., Walling L.L.;
RT   "Localization and post-translational processing of the wound-induced
RT   leucine aminopeptidase proteins of tomato.";
RL   J. Biol. Chem. 271:25880-25887(1996).
RN   [2]
RP   SEQUENCE OF 49-571 FROM N.A.
RC   STRAIN=CV. VF36; TISSUE=Pistil;
RX   MEDLINE=95375233; PubMed=7647301;
RA   Milligan S.B., Gasser C.S.;
RT   "Nature and regulation of pistil-expressed genes in tomato.";
RL   Plant Mol. Biol. 28:691-711(1995).
RN   [3]
RP   SEQUENCE OF 103-571 FROM N.A.
RC   STRAIN=CV. PETO 238R; TISSUE=Leaf;
RX   MEDLINE=94052201; PubMed=8234334;
RA   Pautot V., Holzer F.M., Reisch B., Walling L.L.;
RT   "Leucine aminopeptidase: an inducible component of the defense
RT   response in Lycopersicon esculentum (tomato).";
RL   Proc. Natl. Acad. Sci. U.S.A. 90:9906-9910(1993).
CC   -!- FUNCTION: PRESUMABLY INVOLVED IN THE PROCESSING AND REGULAR
CC       TURNOVER OF INTRACELLULAR PROTEINS.
CC   -!- CATALYTIC ACTIVITY: RELEASE OF AN N-TERMINAL AMINO ACID, XAA-|-
CC       XBB-, IN WHICH XAA IS PREFERABLY LEU, BUT MAY BE OTHER AMINO ACIDS
CC       INCLUDING PRO ALTHOUGH NOT ARG OR LYS, AND XBB MAY BE PRO.
CC   -!- COFACTOR: BINDS TWO ZINC IONS (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: CHLOROPLAST.
CC   -!- INDUCTION: BY WOUNDING.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M17; ALSO KNOWN AS THE
CC       CYTOSOL AMINOPEPTIDASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U50151; AAC49456.1; -.
DR   EMBL; U50152; AAC49457.1; -.
DR   EMBL; U20593; AAA80498.1; -.
DR   HSSP; P00727; 1BPM.
DR   MEROPS; M17.002; -.
DR   InterPro; IPR000819; Peptidase_M17.
DR   Pfam; PF00883; Peptidase_M17; 1.
DR   PRINTS; PR00481; LAMNOPPTDASE.
DR   PROSITE; PS00631; CYTOSOL_AP; 1.
KW   Transit peptide; Chloroplast; Aminopeptidase; Hydrolase; Zinc.
FT   TRANSIT       1     53       CHLOROPLAST (POTENTIAL).
FT   CHAIN        54    571       CHLOROPLAST AMINOPEPTIDASE 1.
FT   DOMAIN      169    174       POLY-ALA.
FT   METAL       342    342       ZINC (2) (BY SIMILARITY).
FT   METAL       347    347       ZINC (1 AND 2) (BY SIMILARITY).
FT   METAL       367    367       ZINC (2) (BY SIMILARITY).
FT   METAL       427    427       ZINC (1) (BY SIMILARITY).
FT   METAL       429    429       ZINC (1 AND 2) (BY SIMILARITY).
FT   ACT_SITE    354    354       POTENTIAL.
FT   ACT_SITE    431    431       POTENTIAL.
FT   VARIANT     358    358       R -> G.
FT   CONFLICT    271    271       P -> N (IN REF. 2).
FT   CONFLICT    315    315       T -> S (IN REF. 3).
FT   CONFLICT    515    515       T -> L (IN CLONE PBLAP2).
SQ   SEQUENCE   571 AA;  60279 MW;  C7A224837E73939D CRC64;
```

```
  Query Match             78.0%;  Score 32;  DB 1;  Length 571;
  Best Local Similarity   83.3%;  Pred. No. 20;
  Matches    5;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 KRIMHC 7
           |||:|| 
Db      48 KRIVHC 53


RESULT  10
AMPL_SOLTU
ID   AMPL_SOLTU     STANDARD;      PRT;   573 AA.
AC   P31427;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   CHLOROPLAST AMINOPEPTIDASE PRECURSOR (EC 3.4.11.1) (LEUCINE
DE   AMINOPEPTIDASE) (LAP) (LEUCYL AMINOPEPTIDASE) (PROLINE AMINOPEPTIDASE)
DE   (EC 3.4.11.5) (PROLYL AMINOPEPTIDASE).
GN   LAP.
OS   Solanum tuberosum (Potato).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC   Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
OX   NCBI_TaxID=4113;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. DESIREE;
RX   MEDLINE=94339796; PubMed=7765119;
RA   Herbers K., Prat S., Willmitzer L.;
RT   "Functional analysis of a leucine aminopeptidase from Solanum
RT   tuberosum L.";
RL   Planta 194:230-240(1994).
RN   [2]
RP   SEQUENCE OF 19-573 FROM N.A.
RC   STRAIN=CV. DESIREE; TISSUE=Leaf;
RX   MEDLINE=93005746; PubMed=1392612;
RA   Hildmann T., Ebneth M., Pena-Cortes H., Sanchez-Serrano J.J.,
RA   Willmitzer L., Prat S.;
RT   "General roles of abscisic and jasmonic acids in gene activation as a
RT   result of mechanical wounding.";
RL   Plant Cell 4:1157-1170(1992).
CC   -!- FUNCTION: PRESUMABLY INVOLVED IN THE PROCESSING AND REGULAR
CC       TURNOVER OF INTRACELLULAR PROTEINS.
CC   -!- CATALYTIC ACTIVITY: RELEASE OF AN N-TERMINAL AMINO ACID, XAA-|-
CC       XBB-, IN WHICH XAA IS PREFERABLY LEU, BUT MAY BE OTHER AMINO ACIDS
CC       INCLUDING PRO ALTHOUGH NOT ARG OR LYS, AND XBB MAY BE PRO.
CC   -!- COFACTOR: BINDS TWO ZINC IONS (BY SIMILARITY).
CC   -!- SUBUNIT: HOMOHEXAMER (PROBABLE).
CC   -!- SUBCELLULAR LOCATION: CHLOROPLAST (BY SIMILARITY).
CC   -!- TISSUE SPECIFICITY: IN TUBERS AND FLORAL BUDS OF UNTREATED PLANTS.
CC       AFTER ABA TREATMENT OR MECHANICAL WOUNDING IS MOSTLY ACCUMULATED
CC       IN LEAVES, TO A LESSER EXTENT IN STEMS, BUT NOT IN ROOTS.
CC   -!- INDUCTION: BY ABSCISIC ACID (ABA), JASMONIC ACID (JA) AND
CC       WOUNDING.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M17; ALSO KNOWN AS THE
CC       CYTOSOL AMINOPEPTIDASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X77015; CAA54314.1; -.
DR   EMBL; X67845; CAA48038.1; -.
DR   PIR; S24769; S24769.
DR   PIR; PQ0470; PQ0470.
DR   HSSP; P00727; 1LAN.
```

```
DR   MEROPS; M17.002; -.
DR   InterPro; IPR000819; Peptidase_M17.
DR   Pfam; PF00883; Peptidase_M17; 1.
DR   PRINTS; PR00481; LAMNOPPTDASE.
DR   PROSITE; PS00631; CYTOSOL_AP; 1.
KW   Transit peptide; Chloroplast; Aminopeptidase; Hydrolase; Zinc.
FT   TRANSIT       1     53       CHLOROPLAST (POTENTIAL).
FT   CHAIN        54    573       CHLOROPLAST AMINOPEPTIDASE.
FT   DOMAIN      169    174       POLY-ALA.
FT   METAL       342    342       ZINC (2) (BY SIMILARITY).
FT   METAL       347    347       ZINC (1 AND 2) (BY SIMILARITY).
FT   METAL       367    367       ZINC (2) (BY SIMILARITY).
FT   METAL       427    427       ZINC (1) (BY SIMILARITY).
FT   METAL       429    429       ZINC (1 AND 2) (BY SIMILARITY).
FT   ACT_SITE    354    354       POTENTIAL.
FT   ACT_SITE    431    431       POTENTIAL.
SQ   SEQUENCE   573 AA;  60122 MW;  3152145A4A7FB291 CRC64;


  Query Match             78.0%;  Score 32;  DB 1;  Length 573;
  Best Local Similarity   83.3%;  Pred. No. 20;
  Matches    5;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 KRIMHC 7
           |||:||
Db      48 KRIVHC 53


RESULT  11
RL44_TRYBB
ID   RL44_TRYBB     STANDARD;      PRT;   105 AA.
AC   P17843;
DT   01-AUG-1990 (Rel. 15, Created)
DT   20-AUG-2001 (Rel. 40, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   60S RIBOSOMAL PROTEIN L44.
GN   RPL44.
OS   Trypanosoma brucei brucei.
OC   Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OX   NCBI_TaxID=5702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=EATRO 1125;
RX   MEDLINE=90251460; PubMed=2339065;
RA   Tebabi P., Halleux S., Pays E.;
RT   "Nucleotide sequence of a full-length cDNA coding for the ribosomal
RT   L44 protein of Trypanosoma brucei.";
RL   Nucleic Acids Res. 18:2809-2809(1990).
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC   -!- SIMILARITY: BELONGS TO THE L44E FAMILY OF RIBOSOMAL PROTEINS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X52122; CAA36367.1; -.
DR   PIR; S10012; R6UT6A.
DR   InterPro; IPR000552; Ribosomal_L44E.
DR   Pfam; PF00935; Ribosomal_L44; 2.
DR   ProDom; PD002841; Ribosomal_L44E; 1.
DR   PROSITE; PS01172; RIBOSOMAL_L44E; 1.
KW   Ribosomal protein.
FT   INIT_MET      0      0       BY SIMILARITY.
SQ   SEQUENCE   105 AA;  12322 MW;  FA19423F109E7819 CRC64;


  Query Match             75.6%;  Score 31;  DB 1;  Length 105;
  Best Local Similarity   71.4%;  Pred. No. 5.9;
  Matches    5;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 KKRIMHC 7
           ||: |||
Db       5 KKKKMHC 11


RESULT  12
Y091_NPVOP
ID   Y091_NPVOP     STANDARD;      PRT;   279 AA.
AC   O10341;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   HYPOTHETICAL 29.3 KDA PROTEIN (ORF92).
OS   Orgyia pseudotsugata multicapsid polyhedrosis virus (OpMNPV).
OC   Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
OC   Nucleopolyhedrovirus.
OX   NCBI_TaxID=164623;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97271300; PubMed=9126251;
RA   Ahrens C.H., Russell R.R., Funk C.J., Evans J., Harwood S.,
RA   Rohrmann G.F.;
RT   "The sequence of the Orgyia pseudotsugata multinucleocapsid nuclear
RT   polyhedrosis virus genome.";
RL   Virology 229:381-399(1997).
CC   -!- SIMILARITY: TO CORRESPONDING ORF IN ACMNPV.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; U75930; AAC59091.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   279 AA;  29289 MW;  6FA4DAA01009DBF0 CRC64;


  Query Match             75.6%;  Score 31;  DB 1;  Length 279;
  Best Local Similarity   66.7%;  Pred. No. 16;
  Matches    4;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 KRIMHC 7
           :|:|||
Db     233 QRVMHC 238


RESULT  13
CYCH_XENLA
ID   CYCH_XENLA     STANDARD;      PRT;   323 AA.
AC   P51947;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   CYCLIN H (MO15-ASSOCIATED PROTEIN) (P36).
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Martinez A.-M.;
RL   Submitted (JAN-1995) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   PARTIAL SEQUENCE.
RC   TISSUE=Oocyte;
RX   MEDLINE=95045408; PubMed=7957080;
RA   Labbe J.-C., Martinez A.-M., Fesquet D., Capony J.-P., Darbon J.-M.,
RA   Derancourt J., Devault A., Morin N., Cavadore J.-C., Doree M.;
```

```
RT   "p40MO15 associates with a p36 subunit and requires both nuclear
RT   translocation and Thr176 phosphorylation to generate cdk-activating
RT   kinase activity in Xenopus oocytes.";
RL   EMBO J. 13:5155-5164(1994).
CC   -!- FUNCTION: SEEMS TO BE A REGULATORY COMPONENT OF CDK-ACTIVATING
CC       KINASE (CAK).
CC   -!- SUBUNIT: ASSOCIATES WITH CDK7 AND MAT1.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR.
CC   -!- SIMILARITY: BELONGS TO THE CYCLIN FAMILY. CYCLIN C SUBFAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; U20505; AAA62236.1; -.
DR   HSSP; P51946; 1JKW.
DR   InterPro; IPR000553; Cyclin.
DR   SMART; SM00385; CYCLIN; 1.
DR   PROSITE; PS00292; CYCLINS; FALSE_NEG.
KW   Cyclin; Cell cycle; Cell division; Nuclear protein.
FT   CONFLICT    266    266       R -> Y (IN REF. 2; AA SEQUENCE).
SQ   SEQUENCE   323 AA;  37600 MW;  14BCDCA000843DC8 CRC64;


  Query Match             75.6%;  Score 31;  DB 1;  Length 323;
  Best Local Similarity   71.4%;  Pred. No. 18;
  Matches    5;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 KKRIMHC 7
           |||: ||
Db     274 KKRLDHC 280


RESULT  14
SUN_HAEIN
ID   SUN_HAEIN      STANDARD;      PRT;   451 AA.
AC   P44788;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   SUN PROTEIN (FMU PROTEIN).
GN   SUN OR FMU OR HI0624.
OS   Haemophilus influenzae.
OC   Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC   Haemophilus.
OX   NCBI_TaxID=727;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=RD / KW20 / ATCC 51907;
RX   MEDLINE=95350630; PubMed=7542800;
RA   Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA   Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA   McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA   Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA   Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA   Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA   Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA   Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA   Venter J.C.;
RT   "Whole-genome random sequencing and assembly of Haemophilus
RT   influenzae Rd.";
RL   Science 269:496-512(1995).
CC   -!- SIMILARITY: BELONGS TO THE SUN (BACTERIAL) / NUCLEOLAR PROTEIN
CC       NOL1/NOP2 (EUKARYOTES) FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; U32745; AAC22284.1; -.
DR   TIGR; HI0624; -.
DR   InterPro; IPR001678; Nol1_Nop2_Sun.
DR   InterPro; IPR000139; NusB.
DR   InterPro; IPR000051; SAM_bind.
DR   Pfam; PF01189; Nol1_Nop2_Sun; 1.
DR   Pfam; PF01029; NusB; 1.
DR   ProDom; PD005242; NusB; 1.
DR   PROSITE; PS01153; NOL1_NOP2_SUN; 1.
KW   Complete proteome.
SQ   SEQUENCE   451 AA;  50597 MW;  D91FAB88FFDE34B0 CRC64;


  Query Match             75.6%;  Score 31;  DB 1;  Length 451;
  Best Local Similarity   71.4%;  Pred. No. 25;
  Matches    5;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 KKRIMHC 7
           | ||:||
Db      88 KTRIVHC 94


RESULT  15
CUL3_HUMAN
ID   CUL3_HUMAN     STANDARD;      PRT;   768 AA.
AC   Q13618; O75415; Q9UBI8; Q9UET7;
DT   01-NOV-1997 (Rel. 35, Created)
DT   20-AUG-2001 (Rel. 40, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   CULLIN HOMOLOG 3 (CUL-3).
GN   CUL3 OR KIAA0617.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98406061; PubMed=9733711;
RA   Du M., Sansores-Garcia L., Zu Z., Wu K.K.;
RT   "Cloning and expression analysis of a novel salicylate suppressible
RT   gene, Hs-CUL-3, a member of cullin/Cdc53 family.";
RL   J. Biol. Chem. 273:24289-24292(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=98403880; PubMed=9734811;
RA   Ishikawa K.-I., Nagase T., Suyama M., Miyajima N., Tanaka A.,
RA   Kotani H., Nomura N., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes. X.
RT   The complete sequences of 100 new cDNA clones from brain which can
RT   code for large proteins in vitro.";
RL   DNA Res. 5:169-176(1998).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Colon carcinoma;
RX   MEDLINE=98326596; PubMed=9663463;
RA   Michel J.J., Xiong Y.;
RT   "Human CUL-1, but not other cullin family members, selectively
RT   interacts with SKP1 to form a complex with SKP2 and cyclin A.";
RL   Cell Growth Differ. 9:435-449(1998).
RN   [4]
RP   SEQUENCE OF 192-768 FROM N.A.
RX   MEDLINE=96279828; PubMed=8681378;
RA   Kipreos E.T., Lander L.E., Wing J.P., He W.W., Hedgecock E.M.;
RT   "cul-1 is required for cell cycle exit in C. elegans and identifies a
RT   novel gene family.";
RL   Cell 85:829-839(1996).
RN   [5]
RP   SEQUENCE OF 426-768 FROM N.A.
```

```
RC   TISSUE=Brain;
RA   Yu W., Sarginson J., Gibbs R.A.;
RL   Submitted (MAR-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: BELONGS TO THE CULLIN FAMILY.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; AF064087; AAC36304.1; -.
DR   EMBL; AB014517; BAA31592.1; -.
DR   EMBL; AF062537; AAC36682.1; -.
DR   EMBL; U58089; AAC50546.1; -.
DR   EMBL; AF052147; AAC28621.1; -.
DR   MIM; 603136; -.
DR   InterPro; IPR001373; Cullin.
DR   InterPro; IPR002119; Histone_H2A.
DR   Pfam; PF00888; Cullin; 1.
DR   ProDom; PD000565; Histone_H2A; 1.
DR   SMART; SM00182; CULLIN; 1.
DR   PROSITE; PS01256; CULLIN_1; 1.
DR   PROSITE; PS50069; CULLIN_2; 1.
FT   CONFLICT     13     13       D -> G (IN REF. 3).
FT   CONFLICT    159    179       DHLRQTLLDMIARERKGEVVD -> GSSTANSIGYDCKRAE
FT                                RRSRS (IN REF. 3).
FT   CONFLICT    426    451       DVFERYYKQHLARRLLTNKSVSDDSE -> MYLNVIINNTW
FT                                QGDFSQIKVFLMTLK (IN REF. 5).
SQ   SEQUENCE   768 AA;  88930 MW;  A1A02022480BF099 CRC64;


  Query Match             75.6%;  Score 31;  DB 1;  Length 768;
  Best Local Similarity   66.7%;  Pred. No. 43;
  Matches    4;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 KRIMHC 7
           :|:|||
Db     246 ERVMHC 251


Search completed: February 12, 2002, 12:04:02
Job time: 798 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: February 12, 2002, 11:51:39 ; Search time 55.4 Seconds
(without alignments)
9.625 Million cell updates/sec

Title: US-09-606-129A-18
Perfect score: 41
Sequence: 1 KKRIMHC 7

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_68:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	41	100.0	295	2	A42268	biliverdin reducta
2	41	100.0	633	2	T04537	hypothetical prote
3	38	92.7	296	2	G02066	biliverdin reducta
4	38	92.7	296	2	S62624	biliverdin reducta
5	36	87.8	512	2	H86206	hypothetical prote
6	34	82.9	401	2	F85015	probable phosphati
7	34	82.9	431	2	T01723	1-phosphatidylinos
8	33	80.5	189	2	A38117	hypothetical prote
9	33	80.5	508	2	T02486	hypothetical prote
10	33	80.5	925	2	S50490	hypothetical prote
11	32	78.0	107	2	B64845	hypothetical prote
12	32	78.0	107	2	D85651	hypothetical prote
13	32	78.0	140	2	I49668	binding protein -
14	32	78.0	142	2	JC1365	FK506/rapamycin-bi
15	32	78.0	328	2	B65086	hydrogenase (EC 1.
16	32	78.0	328	2	A85959	hydrogenase-2 smal
17	32	78.0	377	2	A82213	methylcitrate synt
18	32	78.0	497	2	C86463	hypothetical prote
19	32	78.0	516	2	T23827	hypothetical prote
20	32	78.0	566	2	T07850	leucyl aminopeptid
21	32	78.0	571	2	T07849	leucyl aminopeptid
22	32	78.0	571	2	T07047	leucyl aminopeptid
23	32	78.0	573	1	S41376	leucyl aminopeptid
24	31	75.6	106	1	R6UT6A	ribosomal protein
25	31	75.6	137	2	T46904	hypothetical prote
26	31	75.6	144	2	H41700	C6 protein - rabbi
27	31	75.6	152	2	T32784	hypothetical prote
28	31	75.6	279	2	T10361	hypothetical prote
29	31	75.6	321	2	T48373	hypothetical prote
30	31	75.6	383	2	S51651	cyclin delta-2 - A
31	31	75.6	451	1	F64155	hypothetical prote
32	31	75.6	476	2	T17330	hypothetical prote
33	31	75.6	504	2	JC4775	interferon-induced
34	31	75.6	504	2	A56534	interferon-induced
35	31	75.6	861	2	T41945	primase - human he
36	31	75.6	905	2	B71562	probable oxoglutar
37	31	75.6	905	2	H81714	2-oxoglutarate deh
38	31	75.6	1687	2	T30176	EGF repeat transme
39	30	73.2	108	2	E61180	protein-tyrosine-p
40	30	73.2	108	2	A64511	hypothetical prote
41	30	73.2	269	2	C71255	conserved hypothet
42	30	73.2	370	2	D84464	hypothetical prote
43	30	73.2	457	2	T21344	hypothetical prote
44	30	73.2	508	2	T22440	hypothetical prote
45	30	73.2	522	2	E86453	CDS protein F9L11.

ALIGNMENTS

RESULT 1
A42268
biliverdin reductase (EC 1.3.1.24) - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 05-Nov-1999
C;Accession: A42268
R;Fakhrai, H.; Maines, M.D.
J. Biol. Chem. 267, 4023-4029, 1992
A;Title: Expression and characterization of a cDNA for rat kidney biliverdin reduct
A;Reference number: A42268; MUID:92156147
A;Accession: A42268
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: nucleic acid; protein
A;Residues: 1-295 <FAK>
A;Cross-references: GB:M81681; NID:g203177; PIDN:AAA40830.1; PID:g203178
A;Experimental source: kidney
A;Note: sequence extracted from NCBI backbone (NCBIP:82800)
C;Keywords: liver; oxidoreductase

Query Match 100.0%; Score 41; DB 2; Length 295;
Best Local Similarity 100.0%; Pred. No. 0.36;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KKRIMHC 7
|||||||
Db 274 KKRIMHC 280

RESULT 2
T04537
hypothetical protein F28J12.70 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 14-May-1999
C;Accession: T04537
R;Bevan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Duesterhoeft, A.; Ban
submitted to the Protein Sequence Database, February 1998
A;Reference number: Z15377
A;Accession: T04537
A;Molecule type: DNA
A;Residues: 1-633 <BEV>
A;Cross-references: EMBL:AL021710
A;Experimental source: cultivar Columbia; BAC clone F28J12
C;Genetics:
A;Map position: 4
A;Introns: 281/3; 303/3; 442/1; 614/3
A;Note: F28J12.70

Query Match 100.0%; Score 41; DB 2; Length 633;
Best Local Similarity 100.0%; Pred. No. 0.73;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 KKRIMHC 7
          |||||||
Db    591 KKRIMHC 597
```

```
RESULT   3
G02066
biliverdin reductase (EC 1.3.1.24) - human
N;Alternate names: biliverdin IX-alpha reductase
C;Species: Homo sapiens (man)
C;Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 17-Jul-1998
C;Accession: G02066
R;Komuro, A.; Tobe, T.; Nakano, Y.; Yamaguchi, T.; Tomita, M.
submitted to the EMBL Data Library, August 1995
A;Reference number: H00768
A;Accession: G02066
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-296 <KOM>
A;Cross-references: EMBL:U34877; NID:g1143231; PID:g1143232
C;Keywords: oxidoreductase
```

```
  Query Match             92.7%;  Score 38;  DB 2;  Length 296;
  Best Local Similarity   85.7%;  Pred. No. 1.5;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 KKRIMHC 7
          ||||:||
Db    275 KKRILHC 281
```

```
RESULT   4
S62624
biliverdin reductase (EC 1.3.1.24) - human
N;Alternate names: biliverdin IX-alpha reductase
C;Species: Homo sapiens (man)
C;Date: 28-Oct-1996 #sequence_revision 09-May-1997 #text_change 21-Jul-2000
C;Accession: S62624; S62622; S29736
R;Maines, M.D.; Polevoda, B.V.; Huang, T.J.; McCoubrey Jr., W.K.
Eur. J. Biochem. 235, 372-381, 1996
A;Title: Human biliverdin IX-alpha reductase is a zinc-metalloprotein. Characterization
A;Reference number: S62622; MUID:96202961
A;Accession: S62624
A;Molecule type: mRNA
A;Residues: 1-296 <MAI>
A;Cross-references: EMBL:X93086; NID:g1246748; PIDN:CAA63635.1; PID:g1246749
A;Accession: S62622
A;Molecule type: protein
A;Residues: 3-24,'X',26-27,'X',29-36;48-74;228-234;235-248 <MAF>
R;Maines, M.D.; Trakshel, G.M.
Arch. Biochem. Biophys. 300, 320-326, 1993
A;Title: Purification and characterization of human biliverdin reductase.
A;Reference number: S29736; MUID:93143333
A;Accession: S29736
A;Molecule type: protein
A;Residues: 3-24,'X',26-27,'X',29-36;48-74;228-234;235-248 <MAW>
A;Note: the sequence of peptide 1 from page 323 seems not to belong to this protein
C;Genetics:
A;Gene: BVR
C;Keywords: oxidoreductase
F;3-296/Product: biliverdin reductase IX-alpha #status experimental <MAT>
```

```
  Query Match             92.7%;  Score 38;  DB 2;  Length 296;
  Best Local Similarity   85.7%;  Pred. No. 1.5;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 KKRIMHC 7
          ||||:||
Db    275 KKRILHC 281
```

```
RESULT   5
H86206
hypothetical protein [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C;Accession: H86206
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Al
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Ki
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marz
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tal
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719
A;Accession: H86206
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-512 <STO>
A;Cross-references: GB:AE005172; NID:g8954028; PIDN:AAF82202.1; GSPDB:GN00141
C;Genetics:
A;Map position: 1
```

```
  Query Match             87.8%;  Score 36;  DB 2;  Length 512;
  Best Local Similarity   71.4%;  Pred. No. 6.6;
  Matches    5;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 KKRIMHC 7
          |||::||
Db    236 KKRLLHC 242
```

```
RESULT   6
F85015
probable phosphatidylinositol kinase [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 16-Feb-2001
C;Accession: F85015
R;anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold :
Nature 402, 769-777, 1999
A;Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.
A;Reference number: A85001; MUID:20083488
A;Accession: F85015
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-401 <STO>
A;Cross-references: GB:NC_001268; NID:g7267616; PIDN:CAB80928.1; GSPDB:GN00140
C;Genetics:
A;Gene: AT4g01190
A;Map position: 4
```

```
  Query Match             82.9%;  Score 34;  DB 2;  Length 401;
  Best Local Similarity   85.7%;  Pred. No. 14;
  Matches    6;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 KKRIMHC 7
          |||| ||
Db    353 KKRIEHC 359
```

```
RESULT   7
T01723
1-phosphatidylinositol-4-phosphate 5-kinase type II homolog - Arabidopsis thaliana
N;Alternate names: protein A_IG002N01.9
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 19-Feb-1999 #sequence_revision 19-Feb-1999 #text_change 24-Mar-1999
C;Accession: T01723
R;Scheet, P.; Maggi, L.
```

submitted to the EMBL Data Library, June 1997
A;Description: The sequence of A. thaliana IG002N01.
A;Reference number: Z14407
A;Accession: T01723
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-431 <SCH>
A;Cross-references: EMBL:AF007269; NID:g2191126; PID:g2191143
A;Experimental source: cultivar Columbia
C;Genetics:
A;Map position: 4
A;Introns: 40/2; 94/3; 161/3; 224/2; 255/1; 271/1; 303/1; 339/2
A;Note: A_IG002N01.9

Query Match 82.9%; Score 34; DB 2; Length 431;
Best Local Similarity 85.7%; Pred. No. 15;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 KKRIMHC 7
||||.||
Db 383 KKRIEHC 389

RESULT 8
A38117
hypothetical protein 1 - Anabaena sp. insertion sequence IS895
C;Species: Anabaena sp.
C;Date: 24-Jul-1992 #sequence_revision 24-Jul-1992 #text_change 15-Oct-1999
C;Accession: A38117
R;Alam, J.; Vrba, J.M.; Cai, Y.; Martin, J.A.; Weislo, L.J.; Curtis, S.E.
J. Bacteriol. 173, 5778-5783, 1991
A;Title: Characterization of the IS895 family of insertion sequences from the cyanobacte
A;Reference number: A38117; MUID:91358370
A;Accession: A38117
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-189 <ALA>
A;Cross-references: GB:M67475; NID:g142026; PIDN:AAA98138.1; PID:g142027
A;Experimental source: strain PCC 7120
C;Genetics:
A;Mobile element: insertion sequence IS895

Query Match 80.5%; Score 33; DB 2; Length 189;
Best Local Similarity 71.4%; Pred. No. 11;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 KKRIMHC 7
|||:.||
Db 171 KKRLKHC 177

RESULT 9
T02486
hypothetical protein At2g29990 [imported] - Arabidopsis thaliana
N;Alternate names: hypothetical protein F23F1.9
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 16-Feb-2001
C;Accession: T02486; B84703
R;Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kaul
submitted to the EMBL Data Library, August 1998
A;Description: Arabidopsis thaliana chromosome II BAC F23F1 genomic sequence.
A;Reference number: Z14675
A;Accession: T02486
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-508 <ROU>
A;Cross-references: EMBL:AC004680; NID:g3420043; PID:g3420052
A;Experimental source: cultivar Columbia
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J

Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487
A;Accession: B84703
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-508 <STO>
A;Cross-references: GB:AE002093; NID:g3420052; PIDN:AAC31853.1; GSPDB:GN00139
C;Genetics:
A;Gene: At2g29990; F23F1.9
A;Map position: 2
A;Introns: 158/3; 230/1; 283/3; 305/3; 360/2; 398/3; 458/3
C;Superfamily: NADH dehydrogenase

Query Match 80.5%; Score 33; DB 2; Length 508;
Best Local Similarity 57.1%; Pred. No. 28;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KKRIMHC 7
|:|::||
Db 236 KRRLLHC 242

RESULT 10
S50490
hypothetical protein YER032w - yeast (Saccharomyces cerevisiae)
C;Species: Saccharomyces cerevisiae
C;Date: 28-May-1993 #sequence_revision 24-Feb-1995 #text_change 23-Mar-2001
C;Accession: S50490
R;Dietrich, F.S.
submitted to the EMBL Data Library, December 1994
A;Description: The sequence of S. cerevisiae cosmids 9537, 9581, 9495, 9867, and 1a
A;Reference number: S50433
A;Accession: S50490
A;Molecule type: DNA
A;Residues: 1-925 <DIE>
A;Cross-references: EMBL:U18778; NID:g603592; PIDN:AAB64565.1; PID:g603624; MIPS:YE
C;Genetics:
A;Gene: SGD:FIR1
A;Cross-references: SGD:S0000834; MIPS:YER032w
A;Map position: 5R

Query Match 80.5%; Score 33; DB 2; Length 925;
Best Local Similarity 71.4%; Pred. No. 48;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 KKRIMHC 7
|||:.||
Db 841 KKRLSHC 847

RESULT 11
B64845
hypothetical protein b1028 - Escherichia coli
C;Species: Escherichia coli
C;Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 18-Aug-2000
C;Accession: B64845
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A;Title: The complete genome sequence of Escherichia coli K-12.
A;Reference number: A64720; MUID:97426617
A;Accession: B64845
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-107 <BLAT>
A;Cross-references: GB:AE000205; GB:U00096; NID:g1787265; PIDN:AAC74113.1; PID:g17
A;Experimental source: strain K-12, substrain MG1655
C;Superfamily: Escherichia coli hypothetical protein b1028

```
  Query Match             78.0%;  Score 32;  DB 2;  Length 107;
  Best Local Similarity   57.1%;  Pred. No. 11;
  Matches    4;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 KKRIMHC 7
          ::|:|||
Db     90 QQRVMHC 96


RESULT  12
D85651
hypothetical protein Z1557 [imported] - Escherichia coli (strain O157:H7)
C;Species: Escherichia coli
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Mar-2001
C;Accession: D85651
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A;Reference number: A85480; MUID:21074935; PMID:11206551
A;Accession: D85651
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-107 <STO>
A;Cross-references: GB:AE005174; NID:g12514431; PIDN:AAG55672.1; GSPDB:GN00145; UWGP:Z15
A;Experimental source: strain O157:H7, substrain EDL933
C;Genetics:
A;Gene: Z1557
C;Superfamily: Escherichia coli hypothetical protein b1028


  Query Match             78.0%;  Score 32;  DB 2;  Length 107;
  Best Local Similarity   57.1%;  Pred. No. 11;
  Matches    4;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 KKRIMHC 7
          ::|:|||
Db     90 QQRVMHC 96


RESULT  13
I49668
binding protein - mouse
C;Species: Mus musculus (house mouse)
C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 16-Jul-1999
C;Accession: I49668
R;Hendrickson, B.A.; Zhang, W.; Craig, R.J.; Jin, Y.
Gene 134, 271-275, 1993
A;Title: Structural organization of the genes encoding human and murine FK506-binding pr
A;Reference number: I49668; MUID:94085790
A;Accession: I49668
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-140 <RES>
A;Cross-references: GB:M77831; NID:g433782; PIDN:AAA37631.1; PID:g433783
C;Genetics:
A;Gene: Fkbp13
A;Introns: 55/3; 93/2; 109/1; 121/1
C;Superfamily: BKBP-type peptidylprolyl isomerase; BKBP-type peptidylprolyl isomerase ho
F;47-94/Domain: BKBP-type peptidylprolyl isomerase homology <PPI>


  Query Match             78.0%;  Score 32;  DB 2;  Length 140;
  Best Local Similarity   71.4%;  Pred. No. 14;
  Matches    5;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 KKRIMHC 7
          |||: ||
Db     34 KKRVDHC 40


RESULT  14
JC1365
```

```
FK506/rapamycin-binding protein FKBP13 precursor - human
C;Species: Homo sapiens (man)
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 07-Aug-1998
C;Accession: JC1365; A39602
R;DiLella, A.G.; Hawkins, A.; Craig, R.J.; Schreiber, S.L.; Griffin, C.A.
Biochem. Biophys. Res. Commun. 189, 819-823, 1992
A;Title: Chromosomal band assignments of the genes encoding human FKBP12 and FKBP13
A;Reference number: JC1365; MUID:93112052
A;Accession: JC1365
A;Molecule type: DNA
A;Residues: 1-142 <DIL>
R;Jin, Y.J.; Albers, M.W.; Lane, W.S.; Bierer, B.E.; Schreiber, S.L.; Burakoff, S.J
Proc. Natl. Acad. Sci. U.S.A. 88, 6677-6681, 1991
A;Title: Molecular cloning of a membrane-associated human FK506- and rapamycin-bind
A;Reference number: A39602; MUID:91319747
A;Accession: A39602
A;Molecule type: mRNA
A;Residues: 1-20,'S';23-142 <JIN>
A;Cross-references: GB:M65128
C;Genetics:
A;Gene: GDB:FKBP2
A;Cross-references: GDB:133728; OMIM:186946
A;Map position: 11q13.1-11q13.3
A;Introns: 57/3; 95/2; 111/1; 123/1
C;Superfamily: BKBP-type peptidylprolyl isomerase; BKBP-type peptidylprolyl isomera
C;Keywords: immunoregulation
F;1-22/Domain: signal sequence #status predicted <SIG>
F;23-142/Product: FK506/rapamycin-binding protein FKBP13 #status predicted <MAT>
F;49-96/Domain: BKBP-type peptidylprolyl isomerase homology <PPI>


  Query Match             78.0%;  Score 32;  DB 2;  Length 142;
  Best Local Similarity   71.4%;  Pred. No. 14;
  Matches    5;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 KKRIMHC 7
          |||: ||
Db     36 KKRVDHC 42


RESULT  15
B65086
hydrogenase (EC 1.18.99.1) 2 small chain - Escherichia coli
C;Species: Escherichia coli
C;Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 17-Mar-2000
C;Accession: B65086; A55516
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A;Title: The complete genome sequence of Escherichia coli K-12.
A;Reference number: A64720; MUID:97426617
A;Accession: B65086
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-328 <BLAT>
A;Cross-references: GB:AE000382; GB:U00096; NID:g2367182; PIDN:AAC76032.1; PID:g178
A;Experimental source: strain K-12, substrain MG1655
R;Menon, N.K.; Chatelus, C.Y.; Dervartanian, M.; Wendt, J.C.; Shanmugam, K.T.; Peck
J. Bacteriol. 176, 4416-4423, 1994
A;Title: Cloning, sequencing, and mutational analysis of the hyb operon encoding Es
A;Reference number: A55516; MUID:94292472
A;Accession: A55516
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 'V',2-328 <MEN>
A;Cross-references: GB:U09177
C;Genetics:
A;Gene: hybA
C;Superfamily: unassigned ferredoxin 2[4Fe-4S]-related proteins; ferredoxin 2[4Fe-4
C;Keywords: oxidoreductase
F;105-163/Domain: ferredoxin 2[4Fe-4S] homology <FER>
```

Query Match 78.0%; Score 32; DB 2; Length 328;
Best Local Similarity 71.4%; Pred. No. 30;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```
Qy       1 KKRIMHC 7
           ||: |||
Db     109 KKQCMHC 115
```

Search completed: February 12, 2002, 11:51:40
Job time: 301 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: February 12, 2002, 11:49:43 ; Search time 98.92 Seconds
(without alignments)
5.242 Million cell updates/sec

Title: US-09-606-129A-18
Perfect score: 41
Sequence: 1 KKRIMHC 7

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_1101:*
1: /SIDS8/gcgdata/geneseq/geneseqp/AA1980.DAT:*
2: /SIDS8/gcgdata/geneseq/geneseqp/AA1981.DAT:*
3: /SIDS8/gcgdata/geneseq/geneseqp/AA1982.DAT:*
4: /SIDS8/gcgdata/geneseq/geneseqp/AA1983.DAT:*
5: /SIDS8/gcgdata/geneseq/geneseqp/AA1984.DAT:*
6: /SIDS8/gcgdata/geneseq/geneseqp/AA1985.DAT:*
7: /SIDS8/gcgdata/geneseq/geneseqp/AA1986.DAT:*
8: /SIDS8/gcgdata/geneseq/geneseqp/AA1987.DAT:*
9: /SIDS8/gcgdata/geneseq/geneseqp/AA1988.DAT:*
10: /SIDS8/gcgdata/geneseq/geneseqp/AA1989.DAT:*
11: /SIDS8/gcgdata/geneseq/geneseqp/AA1990.DAT:*
12: /SIDS8/gcgdata/geneseq/geneseqp/AA1991.DAT:*
13: /SIDS8/gcgdata/geneseq/geneseqp/AA1992.DAT:*
14: /SIDS8/gcgdata/geneseq/geneseqp/AA1993.DAT:*
15: /SIDS8/gcgdata/geneseq/geneseqp/AA1994.DAT:*
16: /SIDS8/gcgdata/geneseq/geneseqp/AA1995.DAT:*
17: /SIDS8/gcgdata/geneseq/geneseqp/AA1996.DAT:*
18: /SIDS8/gcgdata/geneseq/geneseqp/AA1997.DAT:*
19: /SIDS8/gcgdata/geneseq/geneseqp/AA1998.DAT:*
20: /SIDS8/gcgdata/geneseq/geneseqp/AA1999.DAT:*
21: /SIDS8/gcgdata/geneseq/geneseqp/AA2000.DAT:*
22: /SIDS8/gcgdata/geneseq/geneseqp/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	36	87.8	285	21	AAG28710	Arabidopsis thalia
2	34	82.9	94	21	AAY97243	Murine INGb1 isofo
3	34	82.9	279	21	AAY97242	Murine P37ING1 pol
4	33	80.5	508	21	AAG48146	Arabidopsis thalia
5	33	80.5	533	21	AAG48145	Arabidopsis thalia
6	33	80.5	564	21	AAG48144	Arabidopsis thalia
7	32	78.0	40	17	AAR93552	Bovine FKBP-13 imm
8	32	78.0	83	22	AAG76114	Human colon cancer
9	32	78.0	94	21	AAY97245	Human INGb1 isofor
10	32	78.0	99	13	AAR28979	Bovine RFKBP. Bos
11	32	78.0	104	21	AAG03758	Human secreted pro
12	32	78.0	128	20	AAY03245	Amino acid sequenc
13	32	78.0	137	22	AAB87656	Bovine mammary tis
14	32	78.0	141	17	AAR93551	Human FKBP-13 immu
15	32	78.0	141	21	AAB53247	Human colon cancer
16	32	78.0	235	22	AAB84700	Amino acid sequenc
17	32	78.0	279	21	AAY97244	Human P37ING1. Ho
18	32	78.0	279	22	AAE06676	Tumour suppressor
19	32	78.0	279	22	AAB84698	Amino acid sequenc
20	32	78.0	294	18	AAW19119	Tumour suppressor
21	32	78.0	294	19	AAW79675	Human p33ING1 poly
22	32	78.0	294	20	AAY03244	Amino acid sequenc
23	32	78.0	294	22	AAB84697	Amino acid sequenc
24	32	78.0	328	22	AAG98906	E. coli growth and
25	32	78.0	528	19	AAW64559	Human protein kina
26	32	78.0	528	22	AAE02011	Human YAK1 (hYAK1)
27	32	78.0	549	21	AAY68781	Amino acid sequenc
28	32	78.0	565	20	AAW96316	Acidic leucine ami
29	32	78.0	571	20	AAW96315	Acidic leucine ami
30	31	75.6	504	18	AAW36140	Bovine P58 protein
31	31	75.6	550	19	AAW71468	Cercospora nicotia
32	31	75.6	550	21	AAY99884	Cercospora nicotia
33	31	75.6	882	22	AAG63551	A human alpha-2 ma
34	31	75.6	887	20	AAY15344	Tumour suppressor
35	31	75.6	887	20	AAY28995	Tumour suppressor
36	31	75.6	899	22	AAG63550	A human alpha-2 ma
37	31	75.6	912	22	AAG63549	A human alpha-2 ma
38	31	75.6	1508	22	AAG63548	A human alpha-2 ma
39	30	73.2	158	21	AAB25386	Pinus radiata cell
40	29	70.7	386	21	AAG21025	Arabidopsis thalia
41	29	70.7	386	21	AAG53432	Arabidopsis thalia
42	29	70.7	391	21	AAB43612	Human cancer assoc
43	29	70.7	396	21	AAY44277	Human nucleic acid
44	29	70.7	396	22	AAM38840	Human polypeptide
45	29	70.7	407	22	AAM40626	Human polypeptide

ALIGNMENTS

RESULT 1
AAG28710
ID AAG28710 standard; Protein; 285 AA.
XX
AC AAG28710;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 34031.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.

PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 87.8%; Score 36; DB 21; Length 285;
Best Local Similarity 71.4%; Pred. No. 14;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 KKRIMHC 7
             |||::||
Db      13 kkrllhc 19
```

RESULT 2
AAY97243
ID AAY97243 standard; Protein; 94 AA.
XX
AC AAY97243;
XX
DT 19-DEC-2000 (first entry)
XX
DE Murine INGb1 isoform N-terminal peptide.
XX
KW p53; tumour; cancer; detection; antibody; hybridisation; exon 1b;
KW ING1; ing1; p33ING1; p37ING1; oncogene; gene therapy; diagnosis;
KW proliferation disorder; transformation; transformed cell; mouse.
XX
OS Mus musculus.
XX
PN WO200046370-A1.
XX
PD 10-AUG-2000.
XX
PF 04-FEB-2000; 2000WO-US02959.
XX
PR 04-FEB-1999; 99US-0118941.
XX
PA (UNII) UNIV ILLINOIS FOUND.
XX
PI Gudkov A, Zeremski M, Gurova KV, Grigorian IA;
XX
DR WPI; 2000-491278/43.
XX
PT Detecting nucleic acid encoding exon 1b of ing1, useful for diagnosing
PT and treating cancer, comprises contacting sample with isolated nucleic
PT acid comprising sequence of exon 1b and detecting hybridized products
XX
PS Claim 6; Fig 7a; 134pp; English.
XX
CC Mutations in or loss of the p53 gene occur in more than 50% of
CC human tumours and tumour cell lines, but functional inactivation of
CC the p53 pathway occurs in a much larger proportion of tumours. In
CC many cases the mechanism of functional inactivation of the p53 gene
CC remains unknown but p53 has been found to act in cooperation with
CC ING1. Functional cooperation between ING1 and p53 suggested that
CC ING1 encoded a tumour suppressor protein that functioned within the
CC p53 pathway. This data suggested a possible role for ING1 in head
CC and neck cancers and chromosomal location of the ING1 placed it
CC within a region that is frequently rearranged in head and neck
CC cancers. Large scale analysis of tumours involving ING1 has not
CC revealed mutations in ING1 nor significant variations in its
CC expression suggesting that ING1 was not a useful gene to study in
CC cancer etiology. However, alternative initiation exons of the ing1
CC gene, each having their own promoter have been discovered.
CC Expression of one promoter (1a) produces a protein identical to
CC ING1. Expression of a second promoter (1b) produces a protein having
CC an identical C-terminal fragment to ING1 but an additional 104
CC N-terminal amino acids. The newly discovered protein has been
CC designated p37ING1 (Wild type: p33ING1). p37ING1 has the
CC characteristics of an oncogene. When overexpressed in cells (even
CC those expressing wild type p53) p37ING1 is able to cause
CC proliferation or transformation of those cells. Thus detecting a
CC nucleic acid encoding exon 1b of ing1 by hybridisation with an
CC isolated nucleic acid having the sequence of exon 1b of ing1
CC or its antisense sequence can identify individuals expressing the
CC oncogenic form of ing1. Novel peptide sequences taken from the 104
CC N-terminal peptide of p37ING1 can also be used to raise antibodies
CC that can also be used in detection methods for the p37ING1 variant.
CC The polypeptides may be useful in gene therapy for treatment of cell
CC proliferation disorders, especially cancers and for diagnosing and
CC studying cancers.
XX
SQ Sequence 94 AA;

Query Match 82.9%; Score 34; DB 21; Length 94;
Best Local Similarity 57.1%; Pred. No. 12;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 KKRIMHC 7
             |:|::||
Db      67 krrvlhc 73
```

RESULT 3
AAY97242
ID AAY97242 standard; Protein; 279 AA.
XX
AC AAY97242;
XX
DT 19-DEC-2000 (first entry)
XX
DE Murine P37ING1 polypeptide.
XX
KW p53; tumour; cancer; detection; antibody; hybridisation; exon 1b;
KW ING1; ing1; p33ING1; p37ING1; oncogene; gene therapy; diagnosis;
KW proliferation disorder; transformation; transformed cell; mouse.
XX
OS Mus musculus.
XX
PN WO200046370-A1.
XX
PD 10-AUG-2000.
XX
PF 04-FEB-2000; 2000WO-US02959.
XX
PR 04-FEB-1999; 99US-0118941.
XX
PA (UNII) UNIV ILLINOIS FOUND.
XX
PI Gudkov A, Zeremski M, Gurova KV, Grigorian IA;
XX
DR WPI; 2000-491278/43.
DR N-PSDB; AAA53790.
XX
PT Detecting nucleic acid encoding exon 1b of ing1, useful for diagnosing
PT and treating cancer, comprises contacting sample with isolated nucleic

PT acid comprising sequence of exon 1b and detecting hybridized products
XX
PS Disclosure; Fig 12; 134pp; English.
XX
CC Mutations in or loss of the p53 gene occur in more than 50% of
CC human tumours and tumour cell lines, but functional inactivation of
CC the p53 pathway occurs in a much larger proportion of tumours. In
CC many cases the mechanism of functional inactivation of the p53 gene
CC remains unknown but p53 has been found to act in cooperation with
CC ING1. Functional cooperation between ING1 and p53 suggested that
CC ING1 encoded a tumour suppressor protein that functioned within the
CC p53 pathway. This data suggested a possible role for ING1 in head
CC and neck cancers and chromosomal location of the ING1 placed it
CC within a region that is frequently rearranged in head and neck
CC cancers. Large scale analysis of tumours involving ING1 has not
CC revealed mutations in ING1 nor significant variations in its
CC expression suggesting that ING1 was not a useful gene to study in
CC cancer etiology. However, alternative initiation exons of the ing1
CC gene, each having their own promoter have been discovered.
CC Expression of one promoter (1a) produces a protein identical to
CC ING1. Expression of a second promoter (1b) produces a protein having
CC an identical C-terminal fragment to ING1 but an additional 104
CC N-terminal amino acids. The newly discovered protein has been
CC designated p37ING1 (Wild type: p33ING1). p37ING1 has the
CC characteristics of an oncogene. When overexpressed in cells (even
CC those expressing wild type p53) p37ING1 is able to cause
CC proliferation or transformation of those cells. Thus detecting a
CC nucleic acid encoding exon 1b of ing1 by hybridisation with an
CC isolated nucleic acid having the sequence of exon 1b of ing1
CC or its antisense sequence can identify individuals expressing the
CC oncogenic form of ing1. Novel peptide sequences taken from the 104
CC N-terminal peptide of p37ING1 can also be used to raise antibodies
CC that can also be used in detection methods for the p37ING1 variant.
CC The polypeptides may be useful in gene therapy for treatment of cell
CC proliferation disorders, especially cancers and for diagnosing and
CC studying cancers.
XX
SQ Sequence 279 AA;

Query Match 82.9%; Score 34; DB 21; Length 279;
Best Local Similarity 57.1%; Pred. No. 34;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 KKRIMHC 7
          |:|::||
Db     67 krrvlhc 73
```

RESULT 4
AAG48146
ID AAG48146 standard; Protein; 508 AA.
XX
AC AAG48146;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 60770.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.

PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 80.5%; Score 33; DB 21; Length 508;
Best Local Similarity 57.1%; Pred. No. 95;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 KKRIMHC 7
             |:|::||
Db     236 krrllhc 242
```

RESULT 5
AAG48145
ID AAG48145 standard; Protein; 533 AA.
XX
AC AAG48145;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 60769.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.

PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 80.5%; Score 33; DB 21; Length 533;
Best Local Similarity 57.1%; Pred. No. 99;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 KKRIMHC 7
          |:|::||
Db    261 krrllhc 267
```

RESULT 6
AAG48144
ID AAG48144 standard; Protein; 564 AA.
XX
AC AAG48144;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 60768.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.

PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 80.5%; Score 33; DB 21; Length 564;
Best Local Similarity 57.1%; Pred. No. 1e+02;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 KKRIMHC 7
             |:|::||
Db     292 krrllhc 298
```

RESULT 7
AAR93552
ID AAR93552 standard; Peptide; 40 AA.
XX
AC AAR93552;
XX
DT 25-JUN-1996 (first entry)
XX
DE Bovine FKBP-13 immunophilin N-terminal peptide.
XX
KW FKBP-13; immunophilin; FK506; rapamycin; rheumatoid arthritis;
KW diabetes; organ transplant; graft versus host disease;
KW immunosuppressant.
XX
OS Bos taurus.
XX
PN US5498597-A.
XX
PD 12-MAR-1996.
XX
PF 17-JAN-1992; 92US-0822966.
XX
PR 17-JAN-1992; 92US-0822966.
XX
PA (DAND) DANA FARBER CANCER INST INC.
PA (HARD) HARVARD COLLEGE.
XX
PI Bierer BE, Burakoff SJ, Schreiber SL;
XX
DR WPI; 1996-159713/16.
XX
PT Purified mammalian FKBP-13 polypeptide capable of binding FK506 -
PT useful for identifying and studying immunosuppressant drugs
XX
PS Disclosure; Column 8; 12pp; English.
XX
CC This sequence encoding the bovine FKBP-13 N-terminal sequence
CC corresponds to the N-terminal sequence of human FKBP-13. FKBP-13
CC may be used for identifying immunosuppressant drugs, and may be
CC used in combination with immunosuppressant drugs for therapeutic
CC purposes in the treatment of autoimmune diseases e.g. rheumatoid
CC arthritis and type-I diabetes, organ transplant and graft versus
CC host disease. The recombinant form of the protein could be
CC potentially smaller and therefore easier to introduce into cells
CC than intact FKBP-13.

```
XX
SQ   Sequence   40 AA;


  Query Match            78.0%;  Score 32;  DB 17;  Length 40;
  Best Local Similarity  71.4%;  Pred. No. 14;
  Matches    5;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 KKRIMHC 7
          |||: ||
Db     14 kkrvdhc 20


RESULT   8
AAG76114
ID   AAG76114 standard; Protein; 83 AA.
XX
AC   AAG76114;
XX
DT   03-SEP-2001  (first entry)
XX
DE   Human colon cancer antigen protein SEQ ID NO:6878.
XX
KW   Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW   colorectal carcinoma; chromosome 11.
XX
OS   Homo sapiens.
XX
PN   WO200122920-A2.
XX
PD   05-APR-2001.
XX
PF   28-SEP-2000; 2000WO-US26524.
XX
PR   29-SEP-1999;   99US-0157137.
PR   03-NOV-1999;   99US-0163280.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Ruben SM,  Barash SC,  Birse CE,  Rosen CA;
XX
DR   WPI; 2001-235357/24.
DR   N-PSDB; AAH35519.
XX
PT   Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT   useful for preventing, diagnosing and/or treating colorectal cancers -
XX
PS   Claim 11; Page 8320-8322; 9803pp; English.
XX
CC   AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
CC   cancer-associated nucleic acid molecules (N) and proteins (P), where
CC   the proteins are collectively known as colon cancer antigens. The colon
CC   cancer antigens have cytostatic activity and can be used in gene
CC   therapy and vaccine production. N and P may be used in the prevention,
CC   diagnosis and treatment of diseases associated with inappropriate P
CC   expression. For example, N and P may be used to treat disorders
CC   associated with decreased expression by rectifying mutations or deletions
CC   in a patient's genome that affect the activity of P by expressing
CC   inactive proteins or to supplement the patients own production of P.
CC   Additionally, N may be used to produce the colon cancer-associated Ps,
CC   by inserting the nucleic acids into a host cell and culturing the cell
CC   to express the proteins. N and P can be used in the prevention, diagnosis
CC   and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
CC   and AAB77789 represent sequences used in the exemplification of the
CC   present invention.
CC   N.B. Pages 666 to 682 and page 7053 of the sequence listing were
CC   missing at time of publication, meaning no sequences are present for
CC   SEQ ID NO:1027 to 1052, 7921 and 7922.
XX
SQ   Sequence   83 AA;


  Query Match            78.0%;  Score 32;  DB 22;  Length 83;
  Best Local Similarity  71.4%;  Pred. No. 27;
  Matches    5;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 KKRIMHC 7
          |||: ||
Db     48 kkrvdhc 54


RESULT   9
AAY97245
ID   AAY97245 standard; Protein; 94 AA.
XX
AC   AAY97245;
XX
DT   19-DEC-2000  (first entry)
XX
DE   Human INGb1 isoform N-terminal peptide.
XX
KW   p53; tumour; cancer; detection; antibody; hybridisation; exon 1b;
KW   ING1; ing1; p33ING1; p37ING1; oncogene; gene therapy; diagnosis;
KW   proliferation disorder; transformation; transformed cell; mouse.
XX
OS   Homo sapiens.
XX
PN   WO200046370-A1.
XX
PD   10-AUG-2000.
XX
PF   04-FEB-2000; 2000WO-US02959.
XX
PR   04-FEB-1999;   99US-0118941.
XX
PA   (UNII ) UNIV ILLINOIS FOUND.
XX
PI   Gudkov A,  Zeremski M,  Gurova KV,  Grigorian IA;
XX
DR   WPI; 2000-491278/43.
XX
PT   Detecting nucleic acid encoding exon 1b of ing1, useful for diagnosing
PT   and treating cancer, comprises contacting sample with isolated nucleic
PT   acid comprising sequence of exon 1b and detecting hybridized products
XX
PS   Claim 26; Figure 7a; 134pp; English.
XX
CC   Mutations in or loss of the p53 gene occur in more than 50% of
CC   human tumours and tumour cell lines, but functional inactivation of
CC   the p53 pathway occurs in a much larger proportion of tumours. In
CC   many cases the mechanism of functional inactivation of the p53 gene
CC   remains unknown but p53 has been found to act in cooperation with
CC   ING1. Functional cooperation between ING1 and p53 suggested that
CC   ING1 encoded a tumour suppressor protein that functioned within the
CC   p53 pathway. This data suggested a possible role for ING1 in head
CC   and neck cancers and chromosomal location of the ING1 placed it
CC   within a region that is frequently rearranged in head and neck
CC   cancers. Large scale analysis of tumours involving ING1 has not
CC   revealed mutations in ING1 nor significant variations in its
CC   expression suggesting that ING1 was not a useful gene to study in
CC   cancer etiology. However, alternative initiation exons of the ing1
CC   gene, each having their own promoter have been discovered.
CC   Expression of one promoter (1a) produces a protein identical to
CC   ING1. Expression of a second promoter (1b) produces a protein having
CC   an identical C-terminal fragment to ING1 but an additional 104
CC   N-terminal amino acids. The newly discovered protein has been
CC   designated p37ING1 (Wild type: p33ING1). p37ING1 has the
CC   characteristics of an oncogene. When overexpressed in cells (even
CC   those expressing wild type p53) p37ING1 is able to cause
CC   proliferation or transformation of those cells. Thus detecting a
CC   nucleic acid encoding exon 1b of ing1 by hybridisation with an
CC   isolated nucleic acid having the sequence of exon 1b of ing1
CC   or its antisense sequence can identify individuals expressing the
CC   oncogenic form of ing1. Novel peptide sequences taken from the 104
CC   N-terminal peptide of p37ING1 can also be used to raise antibodies
CC   that can also be used in detection methods for the p37ING1 variant.
```

```
CC   The polypeptides may be useful in gene therapy for treatment of cell
CC   proliferation disorders, especially cancers and for diagnosing and
CC   studying cancers.
XX
SQ   Sequence   94 AA;


  Query Match             78.0%;  Score 32;  DB 21;  Length 94;
  Best Local Similarity   57.1%;  Pred. No. 30;
  Matches    4;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 KKRIMHC 7
           |:|::||
Db      67 krrmlhc 73


RESULT  10
AAR28979
ID   AAR28979 standard; protein; 99 AA.
XX
AC   AAR28979;
XX
DT   24-MAR-1993  (first entry)
XX
DE   Bovine RFKBP.
XX
KW   Rapamycin; FK506; binding protein; RFKBP; prolyl isomerase;
KW   immunosupressant; cyclosporin A; macrolide; bovine; thymus; bRFKBP;
KW   cis-trans prolyl isomerase activity; FKBP12.
XX
OS   Bos taurus.
XX
PN   WO9219745-A.
XX
PD   12-NOV-1992.
XX
PF   07-MAY-1992;   92WO-US03993.
XX
PR   08-MAY-1991;   91US-0697113.
XX
PA   (VERT-) VERTEX PHARM INC.
XX
PI   Harding MW;
XX
DR   WPI; 1992-398871/48.
XX
PT   New prolyl isomerase and rapamycin FK506 binding protein - useful
PT   for screening potential immunosuppressive cpds.
XX
PS   Disclosure; Fig 1; 30pp; English.
XX
CC   This sequence corresponds to a fragment of a rapamycin FK506 binding
CC   protein (RFKBP).  RFKBP is a prolyl isomerase structurally related to
CC   FK506 which does not bind the immunosupressive cyclosporin A.  RKFBP
CC   binds FK506 and rapamycin with quantitatively significant selectivity.
CC   RKFBP may be used in screening assays to detect new immunosupressants
CC   and to differentiate rapamycin-like cpds. from FK506-like cpds.
CC   Rapamycin is a macrolide which is structurally related to FK506.
CC   This RFKBP has been isolated from bovine thymus (bRFKBP) and was found
CC   to be of low molecular weight, approx. 16,000, and to have cis-trans
CC   prolyl isomerase activity. The N terminal of bRFKBP has been shown to
CC   have over 50% homology to the N terminal of FKBP12.
XX
SQ   Sequence   99 AA;


  Query Match             78.0%;  Score 32;  DB 13;  Length 99;
  Best Local Similarity   71.4%;  Pred. No. 32;
  Matches    5;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 KKRIMHC 7
           |||: ||
Db      14 kkrvdhc 20


RESULT  11
AAG03758
ID   AAG03758 standard; Protein; 104 AA.
XX
AC   AAG03758;
XX
DT   06-OCT-2000  (first entry)
XX
DE   Human secreted protein, SEQ ID NO: 7839.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW   gene therapy; chromosome mapping.
XX
OS   Homo sapiens.
XX
PN   EP1033401-A2.
XX
PD   06-SEP-2000.
XX
PF   21-FEB-2000; 2000EP-0200610.
XX
PR   26-FEB-1999;   99US-0122487.
XX
PA   (GEST ) GENSET.
XX
PI   Dumas Milne Edwards J,  Duclert A,  Giordano J;
XX
DR   WPI; 2000-500381/45.
DR   N-PSDB; AAC03764.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS   Claim 13; SEQ ID 7839; 71pp + CD-ROM; English.
XX
CC   The present sequence is a polypeptide encoded by one of a large number
CC   of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs
CC   were prepared from total human RNAs or polyA+ RNAs derived from 30
CC   different tissues. EST sequences usually correspond mainly to the 3'
CC   untranslated region (UTR) of the mRNA because they are often obtained
CC   from oligo-dT primed cDNA libraries. Such ESTs are not well suited for
CC   isolating cDNA sequences derived from the 5' ends of mRNAs and even in
CC   those cases where longer cDNA sequences have been obtained, the full 5'
CC   UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'
CC   ends and can therefore be used to obtain full length cDNAs and genomic
CC   DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and
CC   chromosome mapping procedures. They are used to obtain upstream
CC   regulatory sequences and to design expression and secretion vectors.
XX
SQ   Sequence   104 AA;


  Query Match             78.0%;  Score 32;  DB 21;  Length 104;
  Best Local Similarity   71.4%;  Pred. No. 33;
  Matches    5;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 KKRIMHC 7
           |||: ||
Db      36 kkrvdhc 42


RESULT  12
AAY03245
ID   AAY03245 standard; Protein; 128 AA.
XX
AC   AAY03245;
XX
DT   26-AUG-1999  (first entry)
XX
DE   Amino acid sequence of the human p28-ING1 protein.
```

```
XX
KW   Human; p33-ING1 protein; growth regulation; apoptosis; DNA damage;
KW   inhibition; anchorage independent growth; cytotoxic drug; cancer;
KW   transcriptional activation; immortal cell line; p28-ING1 protein.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc_Difference 59..269
FT                   /note= "p26-ING1 fragment"
XX
PN   WO9916790-A1.
XX
PD   08-APR-1999.
XX
PF   24-SEP-1998;    98WO-US18179.
XX
PR   14-JAN-1998;    98US-0006783.
PR   26-SEP-1997;    97US-0060138.
XX
PA   (UNII ) UNIV ILLINOIS BOARD OF TRUSTEES.
PA   (UYTE-) UNIV TECHNOLOGIES INT INC.
XX
PI   Garkavtsev I,  Gudkov A,  Riabowol K;
XX
DR   WPI; 1999-263685/22.
XX
PT   Use of p33-ING1 peptides
XX
PS   Example 8; Page 61; 64pp; English.
XX
CC   This is the amino acid sequence of the human p28-ING1 protein,
CC   used in the method of the invention, involving the human p33-ING1
CC   protein. The ING1 gene encodes p33-ING1 which can be used to
CC   modulate the activity of, isolate or detect p53. Expression of the
CC   ING1 and p53 genes in a mammalian cell results in normal growth
CC   regulation anchorage-dependent growth and apoptosis as a response
CC   to irreversible DNA damage and other cellular insult. Inhibition of
CC   expression of either gene results in a loss of cellular growth
CC   control, anchorage independent growth, inhibition of apoptosis
CC   and resistance to radiation and cytotoxic drugs. The p33-ING1 is a
CC   component of the p53 signalling pathway that cooperates with p53 in
CC   negative regulation of cell proliferation by modulating p53 dependent
CC   transcriptional activation.  Biological function of p53 signalling
CC   pathway can therefore be regulated (both enhanced or suppressed) by
CC   modulating p33-ING1 activity. The modulation of p33-ING1 activity can
CC   be used for the stimulation or restoration of the p53 pathway in
CC   anti cancer therapy or for the suppression of the p53 pathway to
CC   defend sensitive tissues from genotoxic stress or for the generation
CC   of immortal cell lines.
XX
SQ   Sequence   128 AA;


  Query Match             78.0%;  Score 32;  DB 20;  Length 128;
  Best Local Similarity   57.1%;  Pred. No. 41;
  Matches    4;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 KKRIMHC 7
           |:|::||
Db      67 krrmlhc 73


RESULT  13
AAB87656
ID   AAB87656 standard; protein; 137 AA.
XX
AC   AAB87656;
XX
DT   15-MAY-2001  (first entry)
XX
DE   Bovine mammary tissue derived protein #47.
XX
KW   Bovine; mammary gland; cancer; tumour; angiogenesis.
XX
OS   Bos taurus.
XX
PN   WO200114553-A1.
XX
PD   01-MAR-2001.
XX
PF   23-AUG-2000; 2000WO-NZ00166.
XX
PR   23-AUG-1999;   99US-0150330.
XX
PA   (GENE-) GENESIS RES & DEV CORP LTD.
PA   (NZPA-) NEW ZEALAND PASTORAL AGRIC RES INST LTD.
XX
PI   Havukkala IJ,  Gleen M,  Grigor MR,  Molenaar AJ;
XX
DR   WPI; 2001-226619/23.
XX
PT   New polypeptides and polynucleotides encoding the polypeptides, which
PT   are expressed in bovine mammary gland tissue, useful for stimulating
PT   mammary gland growth or function, or inducing differentiation of milk
PT   producing cells  -
XX
PS   Claim 11; Page 80; 97pp; English.
XX
CC   The present invention relates to proteins derived from bovine
CC   mammary gland cells. The invention is useful for stimulating
CC   bovine mammary gland cell growth and function, inhibiting the
CC   growth of various mammary gland cancer cells, inhibiting
CC   angiogenesis and vascularization of tumours, or modulating
CC   the growth of blood vessels in a mammal.
XX
SQ   Sequence   137 AA;


  Query Match             78.0%;  Score 32;  DB 22;  Length 137;
  Best Local Similarity   71.4%;  Pred. No. 43;
  Matches    5;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 KKRIMHC 7
             |||: ||
Db      43 kkrvdhc 49


RESULT  14
AAR93551
ID   AAR93551 standard; Protein; 141 AA.
XX
AC   AAR93551;
XX
DT   25-JUN-1996  (first entry)
XX
DE   Human FKBP-13 immunophilin protein.
XX
KW   FKBP-13; immunophilin; FK506; rapamycin; rheumatoid arthritis;
KW   diabetes; organ transplant; graft versus host disease;
KW   immunosuppressant.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..21
FT                   /label= sig_peptide
FT   Misc-difference 22..61
FT                   /note= "corresponds to bovine thymus FKBP-13 N-
FT                          terminal sequence"
FT   Misc-difference 117..120
FT                   /note= "endoplasmic reticulum retention sequence"
XX
PN   US5498597-A.
XX
PD   12-MAR-1996.
```

XX
PF 17-JAN-1992; 92US-0822966.
XX
PR 17-JAN-1992; 92US-0822966.
XX
PA (DAND) DANA FARBER CANCER INST INC.
PA (HARD) HARVARD COLLEGE.
XX
PI Bierer BE, Burakoff SJ, Schreiber SL;
XX
DR WPI; 1996-159713/16.
DR N-PSDB; AAT18037.
XX
PT Purified mammalian FKBP-13 polypeptide capable of binding FK506 -
PT useful for identifying and studying immunosuppressant drugs
XX
PS Claim 1; Fig.1; 12pp; English.
XX
CC The FKBP-13 protein may be used for identifying immunosuppressant
CC drugs, and may be used in combination with immunosuppressant drugs
CC for therapeutic purposes in the treatment of autoimmune diseases e.g.
CC rheumatoid arthritis and type-I diabetes, organ transplant and
CC graft versus host disease. The recombinant form of the protein
CC could be potentially smaller and therefore easier to introduce
CC into cells than intact FKBP-13.
XX
SQ Sequence 141 AA;

Query Match 78.0%; Score 32; DB 17; Length 141;
Best Local Similarity 71.4%; Pred. No. 45;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 KKRIMHC 7
|||: ||
Db 35 kkrvdhc 41

RESULT 15
AAB53247
ID AAB53247 standard; Protein; 141 AA.
XX
AC AAB53247;
XX
DT 09-MAR-2001 (first entry)
XX
DE Human colon cancer antigen protein sequence SEQ ID NO:787.
XX
KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW identification; cytostatic; cardioactive; neuroprotective; vulnerary;
KW immunomodulatory; muscular; gynaecological; gastrointestinal;
KW nephrotropic; antiinfective; antibacterial; gene therapy; wound;
KW neural disorder; immune system disorder; muscular disorder;
KW reproductive disorder; gastrointestinal disorder; renal disorder;
KW infectious disease; cardiovascular disorder.
XX
OS Homo sapiens.
XX
PN WO200055351-A1.
XX
PD 21-SEP-2000.
XX
PF 08-MAR-2000; 2000WO-US05883.
XX
PR 12-MAR-1999; 99US-0124270.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM;
XX
DR WPI; 2000-587534/55.
DR N-PSDB; AAC98004.
XX
PT Colon cancer associated gene sequences, referred to as colon cancer
PT antigens, useful for the treatment, prevention, and diagnosis of colon
PT disorders such as colon cancer -
XX
PS Claim 11; Page 1346; 2104pp; English.
XX
CC AAC97991 to AAC98763 encode the human colon cancer associated proteins,
CC called human colon cancer antigens, given in AAB53234 to AAB54006. The
CC human colon cancer antigens can have cytostatic, cardioactive, muscular;
CC neuroprotective, immunomodulatory, gynaecological, gastrointestinal,
CC vulnerary, nephrotropic, antiinfective and antibacterial activities, and
CC can be used in gene therapy. The colon cancer antigen polynucleotides,
CC proteins and antibodies to the proteins are useful for the prevention,
CC treatment and diagnosis of colon disorders, such as colon cancer. The
CC polynucleotides may be used in diagnostics and research, such as for
CC chromosome identification, and as hybridisation probes. The proteins
CC may also be used to prevent diseases such as neural disorders, immune
CC system disorders, muscular disorders, reproductive disorders,
CC gastrointestinal disorders, wounds, renal disorders, infectious
CC diseases, and cardiovascular disorders. AAC98764 to AAC98772 and
CC AAB54007 represent sequences used in the exemplification of the present
CC invention.
XX
SQ Sequence 141 AA;

Query Match 78.0%; Score 32; DB 21; Length 141;
Best Local Similarity 71.4%; Pred. No. 45;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 KKRIMHC 7
| ||:||
Db 27 knriihc 33

Search completed: February 12, 2002, 11:49:43
Job time: 499 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: February 12, 2002, 12:03:25 ; Search time 94.82 Seconds
(without alignments)
12.341 Million cell updates/sec

Title: US-09-606-129A-19
Perfect score: 46
Sequence: 1 QKLCHQKK 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_17:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	37	80.4	636	12	O90142	O90142 spodoptera
2	37	80.4	1063	12	Q9J844	Q9j844 spodoptera
3	35	76.1	295	11	Q9CY64	Q9cy64 mus musculu
4	35	76.1	303	11	Q9DD21	Q9dd21 mus musculu
5	35	76.1	420	10	O22019	O22019 cyanidiosch
6	34	73.9	208	5	Q94164	Q94164 caenorhabdi
7	34	73.9	322	4	Q9BSM2	Q9bsm2 homo sapien
8	34	73.9	409	11	Q9D3R6	Q9d3r6 mus musculu
9	34	73.9	522	4	Q9H8F4	Q9h8f4 homo sapien
10	34	73.9	547	3	O74308	O74308 schizosacch
11	34	73.9	790	4	Q9NVK9	Q9nvk9 homo sapien
12	34	73.9	1188	5	O96143	O96143 plasmodium
13	34	73.9	1295	3	O13348	O13348 magnaporthe
14	34	73.9	1473	11	Q63625	Q63625 rattus norv
15	34	73.9	1654	4	Q9P1Y6	Q9ply6 homo sapien
16	33	71.7	261	4	Q9H1N6	Q9h1n6 homo sapien
17	33	71.7	263	1	Q9YDU2	Q9ydu2 aeropyrum p
18	33	71.7	266	2	P96301	P96301 alcaligenes
19	33	71.7	273	12	Q9PYW7	Q9pyw7 xestia c-ni
20	33	71.7	445	10	Q9ZPL5	Q9zpl5 nicotiana t
21	33	71.7	540	8	Q9MT29	Q9mt29 solanum tub
22	33	71.7	566	4	Q9H2S0	Q9h2s0 homo sapien
23	33	71.7	588	4	Q9ULU1	Q9ulu1 homo sapien
24	33	71.7	673	11	Q9QVT4	Q9qvt4 mus sp. mpl
25	33	71.7	773	10	O80631	O80631 arabidopsis
26	33	71.7	1247	5	Q17461	Q17461 caenorhabdi
27	32	69.6	87	5	Q9N5N6	Q9n5n6 caenorhabdi
28	32	69.6	163	2	Q9EUG2	Q9eug2 escherichia
29	32	69.6	312	11	Q9CSE2	Q9cse2 mus musculu
30	32	69.6	336	4	Q9H495	Q9h495 homo sapien
31	32	69.6	380	5	Q21866	Q21866 caenorhabdi
32	32	69.6	397	4	Q9H497	Q9h497 homo sapien
33	32	69.6	397	4	Q9H6E7	Q9h6e7 homo sapien
34	32	69.6	422	10	Q9SSP1	Q9ssp1 arabidopsis
35	32	69.6	462	5	Q9U0E3	Q9u0e3 strongyloce
36	32	69.6	465	5	Q9XW94	Q9xw94 caenorhabdi
37	32	69.6	474	10	O80588	O80588 arabidopsis
38	32	69.6	548	5	Q20367	Q20367 caenorhabdi
39	32	69.6	590	2	Q45490	Q45490 bacillus st
40	32	69.6	628	12	O90139	O90139 mamestra br
41	32	69.6	828	11	Q9QZ07	Q9qz07 mus musculu
42	32	69.6	828	11	Q9D6I6	Q9d6i6 mus musculu
43	32	69.6	843	12	Q994E8	Q994e8 porcine ade
44	32	69.6	864	10	Q9C987	Q9c987 arabidopsis
45	32	69.6	920	11	Q9JJK7	Q9jjk7 mesocricetu

ALIGNMENTS

RESULT 1
O90142
ID O90142 PRELIMINARY; PRT; 636 AA.
AC O90142;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE DNA POLYMERASE (FRAGMENT).
GN DPOL.
OS Spodoptera exigua nucleopolyhedrovirus.
OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
OC Nucleopolyhedrovirus.
OX NCBI_TaxID=10454;
RN [1]
RP SEQUENCE FROM N.A.
RA Bulach D.M., Kumar C.A., Zaia A., Liang B., Tribe D.E.;
RT "Group II Nucleopolyhedrovirus Subgroups Revealed by Phylogenetic
RT Analysis of Polyhedrin and DNA Polymerase Gene Sequences.";
RL Submitted (MAY-1998) to the EMBL/GenBank/DDBJ databases.
CC -!- CATALYTIC ACTIVITY: N DEOXYNUCLEOSIDE TRIPHOSPHATE = N
CC PYROPHOSPHATE + DNA(N).
CC -!- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-B FAMILY.
DR EMBL; AF068186; AAC33749.1; -.
DR InterPro; IPR002064; DNA_pol_B.
DR Pfam; PF00136; DNA_pol_B; 2.
DR PRINTS; PR00106; DNAPOLB.
DR SMART; SM00486; POLBc; 1.
KW DNA replication; DNA-binding; DNA-directed DNA polymerase.
FT NON_TER 1 1
FT NON_TER 636 636
SQ SEQUENCE 636 AA; 73836 MW; 9BDB36475B07DCD1 CRC64;

Query Match 80.4%; Score 37; DB 12; Length 636;
Best Local Similarity 85.7%; Pred. No. 12;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 KLCHQKK 8
|||||:
Db 392 KLCHQKQ 398

```
RESULT   2
Q9J844
ID   Q9J844      PRELIMINARY;      PRT;  1063 AA.
AC   Q9J844;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   ORF93 DNA POLYMERASE.
OS   Spodoptera exigua nucleopolyhedrovirus.
OC   Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
OC   Nucleopolyhedrovirus.
OX   NCBI_TaxID=10454;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20036646; PubMed=10567663;
RA   IJkel W.F., van Strien E.A., Heldens J.G., Broer R., Zuidema D.,
RA   Goldbach R.W., Vlak J.M.;
RT   "Sequence and organization of the spodoptera exigua multicapsid
RT   nucleopolyhedrovirus genome.";
RL   J. Gen. Virol. 80:3289-3304(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   IJkel W.F.J., van Strien E.A., Heldens J.G.M., Broer R., Zuidema D.,
RA   Goldbach R.W., Vlak J.M.;
RL   Submitted (JUL-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- CATALYTIC ACTIVITY: N DEOXYNUCLEOSIDE TRIPHOSPHATE = N
CC       PYROPHOSPHATE + DNA(N).
CC   -!- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-B FAMILY.
DR   EMBL; AF169823; AAF33622.1; -.
DR   InterPro; IPR002064; DNA_pol_B.
DR   Pfam; PF00136; DNA_pol_B; 3.
DR   PRINTS; PR00106; DNAPOLB.
DR   SMART; SM00486; POLBc; 1.
KW   DNA replication; DNA-binding; DNA-directed DNA polymerase.
SQ   SEQUENCE   1063 AA;  123108 MW;  471603FAA92B9A10 CRC64;


  Query Match             80.4%;  Score 37;  DB 12;  Length 1063;
  Best Local Similarity   85.7%;  Pred. No. 18;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 KLCHQKK 8
           ||||||:
Db     557 KLCHQKQ 563


RESULT   3
Q9CY64
ID   Q9CY64      PRELIMINARY;      PRT;   295 AA.
AC   Q9CY64;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   2500001N03RIK PROTEIN.
GN   2500001N03RIK.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=EMBRYONIC LIVER;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
DR   EMBL; AK010847; BAB27219.1; -.
DR   MGD; MGI:1917355; 2500001N03Rik.
DR   InterPro; IPR000683; GFO_IDH_MocA.
DR   Pfam; PF01408; GFO_IDH_MocA; 1.
SQ   SEQUENCE   295 AA;  33524 MW;  F2E1682BD77032A4 CRC64;


  Query Match             76.1%;  Score 35;  DB 11;  Length 295;
  Best Local Similarity   62.5%;  Pred. No. 18;
  Matches    5;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 QKLCHQKK 8
             |:|||:|:
Db     288 QRLCHRKQ 295


RESULT   4
Q9DD21
ID   Q9DD21      PRELIMINARY;      PRT;   303 AA.
AC   Q9DD21;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   0610006A11RIK PROTEIN.
GN   0610006A11RIK.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=KIDNEY;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
DR   EMBL; AK002231; BAB21950.1; -.
DR   MGD; MGI:1915580; 0610006A11Rik.
DR   InterPro; IPR000683; GFO_IDH_MocA.
DR   Pfam; PF01408; GFO_IDH_MocA; 1.
SQ   SEQUENCE   303 AA;  34491 MW;  52D8A3B02EE956EB CRC64;


  Query Match             76.1%;  Score 35;  DB 11;  Length 303;
  Best Local Similarity   62.5%;  Pred. No. 18;
  Matches    5;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy      1 QKLCHQKK 8
          |:|||:|:
Db    288 QRLCHRKQ 295


RESULT   5
O22019
ID   O22019      PRELIMINARY;      PRT;   420 AA.
AC   O22019;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   ORF420.
OS   Cyanidioschyzon merolae.
OC   Eukaryota; Rhodophyta; Bangiophyceae; Cyanidioschyzon.
OX   NCBI_TaxID=45157;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Ohta N.;
RL   J. Plant Res. 110:235-245(1997).
DR   EMBL; D63675; BAA22815.1; -.
DR   Mendel; 23995; Cyame;3036;23995.
DR   InterPro; IPR000178; IF2.
DR   InterPro; IPR001950; SUI1.
DR   Pfam; PF02131; IF2; 1.
DR   ProDom; PD186100; IF2; 1.
DR   PROSITE; PS01118; SUI1_1; UNKNOWN_1.
SQ   SEQUENCE   420 AA;  47691 MW;  A6CAE107B24B4E19 CRC64;


  Query Match             76.1%;  Score 35;  DB 10;  Length 420;
  Best Local Similarity   85.7%;  Pred. No. 23;
  Matches    6;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 QKLCHQK 7
          |||||  |
Db    236 QKLCHPK 242


RESULT   6
Q94164
ID   Q94164      PRELIMINARY;      PRT;   208 AA.
AC   Q94164;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   SIMILARITY TO BPTI/KUNITZ INHIBITOR DOMAIN.
GN   C10G8.2.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE=94150718; PubMed=7906398;
RA   Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA   Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA   Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA   Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA   Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA   Lightning J., Lloyd C., Mcmurray A., Mortimore B., O'Callaghan M.,
RA   Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA   Smaldon N., Smith A., Sonnhammer E., Staden R., Sulston J.,
RA   Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA   Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Blanchard M., Bradshaw H.;
RL   Submitted (OCT-1996) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Waterston R.;
RL   Submitted (SEP-1996) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: TO PANCREATIC TRYPSIN INHIBITOR (KUNITZ) DOMAIN.
DR   EMBL; U70857; AAB09170.1; -.
DR   InterPro; IPR002223; Kunitz_BPTI.
DR   Pfam; PF00014; Kunitz_BPTI; 1.
DR   SMART; SM00131; KU; 1.
DR   PROSITE; PS00280; BPTI_KUNITZ_1; 1.
DR   PROSITE; PS50279; BPTI_KUNITZ_2; 1.
KW   Serine protease inhibitor.
SQ   SEQUENCE   208 AA;  24008 MW;  779AB4A8948E67B0 CRC64;


  Query Match             73.9%;  Score 34;  DB 5;  Length 208;
  Best Local Similarity   83.3%;  Pred. No. 21;
  Matches    5;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      3 LCHQKK 8
          :|||||
Db    172 MCHQKK 177


RESULT   7
Q9BSM2
ID   Q9BSM2      PRELIMINARY;      PRT;   322 AA.
AC   Q9BSM2;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   UNKNOWN (PROTEIN FOR IMAGE:3619689) (FRAGMENT).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LYMPHOMA;
RA   Strausberg R.;
RL   Submitted (MAR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC004950; AAH04950.1; -.
FT   NON_TER       1       1
SQ   SEQUENCE   322 AA;  34916 MW;  33C65E86B8F9D761 CRC64;


  Query Match             73.9%;  Score 34;  DB 4;  Length 322;
  Best Local Similarity   71.4%;  Pred. No. 29;
  Matches    5;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 QKLCHQK 7
          ||:|| |
Db    274 QKICHSK 280


RESULT   8
Q9D3R6
ID   Q9D3R6      PRELIMINARY;      PRT;   409 AA.
AC   Q9D3R6;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   4933439B08RIK PROTEIN.
GN   4933439B08RIK.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
```

```
RC   STRAIN=C57BL/6J; TISSUE=TESTIS;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
DR   EMBL; AK017114; BAB30604.1; -.
DR   MGD; MGI:1918456; 4933439B08Rik.
DR   InterPro; IPR003593; AAA.
DR   InterPro; IPR003959; AAA_subfam.
DR   Pfam; PF00004; AAA; 1.
DR   SMART; SM00382; AAA; 1.
SQ   SEQUENCE   409 AA;  46131 MW;  9600B20008DC9749 CRC64;


  Query Match             73.9%;  Score 34;  DB 11;  Length 409;
  Best Local Similarity   83.3%;  Pred. No. 35;
  Matches    5;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 KLCHQK 7
           |:||||
Db     122 KICHQK 127


RESULT   9
Q9H8F4
ID   Q9H8F4      PRELIMINARY;      PRT;   522 AA.
AC   Q9H8F4;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   CDNA FLJ13679 FIS, CLONE PLACE2000006.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=PLACENTA;
RA   Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA   Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,
RA   Tanase T., Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K.,
RA   Arita M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,
RA   Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y., Oshima A.;
RT   "NEDO human cDNA sequencing project.";
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK023741; BAB14663.1; -.
DR   InterPro; IPR001849; PH.
DR   InterPro; IPR000219; RhoGEF.
DR   Pfam; PF00169; PH; 1.
DR   Pfam; PF00621; RhoGEF; 1.
DR   SMART; SM00233; PH; 1.
DR   SMART; SM00325; RhoGEF; 1.
DR   PROSITE; PS50003; PH_DOMAIN; 1.
SQ   SEQUENCE   522 AA;  59222 MW;  F01A6BF70D2920F9 CRC64;


  Query Match             73.9%;  Score 34;  DB 4;  Length 522;
  Best Local Similarity   62.5%;  Pred. No. 42;
  Matches    5;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 QKLCHQKK 8
           |:||||::
Db     157 QELCHQQE 164


RESULT   10
O74308
ID   O74308      PRELIMINARY;      PRT;   547 AA.
AC   O74308;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   ZINC-FINGER PROTEIN.
GN   SPBC15D4.02.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972H-;
RA   Lyne M., Rajandream M.A., Barrell B.G., Lucas M., Gaillardin C.;
RL   Submitted (AUG-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC   -!- SIMILARITY: CONTAINS A ZN(2)-CYS(6), FUNGAL-TYPE BINUCLEAR CLUSTER
CC       DOMAIN.
DR   EMBL; AL031349; CAA20477.1; -.
DR   HSSP; P25502; 1ZME.
DR   InterPro; IPR001138; Zn2_CY6_fungal.
DR   Pfam; PF00172; Zn_clus; 1.
DR   SMART; SM00066; GAL4; 1.
DR   PROSITE; PS00463; ZN2_CY6_FUNGAL_1; 1.
DR   PROSITE; PS50048; ZN2_CY6_FUNGAL_2; 1.
KW   DNA-binding; Metal-binding; Nuclear protein; Transcription regulation;
KW   Zinc; Zinc-finger.
SQ   SEQUENCE   547 AA;  59641 MW;  A65FD7D039B6CD18 CRC64;


  Query Match             73.9%;  Score 34;  DB 3;  Length 547;
  Best Local Similarity   71.4%;  Pred. No. 43;
  Matches    5;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 QKLCHQK 7
           :||||:|
Db     427 RKLCHEK 433


RESULT   11
Q9NVK9
ID   Q9NVK9      PRELIMINARY;      PRT;   790 AA.
AC   Q9NVK9;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   CDNA FLJ10665 FIS, CLONE NT2RP2006200.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA   Nishikawa T., Nagai K., Sugano S., Ishibashi T., Fujimori K.,
RA   Tanai H., Kimata M., Watanabe M., Hiraoka S., Ishii S., Kawai Y.,
RA   Saito K., Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K.,
RA   Masuho Y., Kanehori K.;
RT   "NEDO human cDNA sequencing project.";
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
```

```
DR   EMBL; AK001527; BAA91741.1; -.
DR   InterPro; IPR001849; PH.
DR   InterPro; IPR000219; RhoGEF.
DR   Pfam; PF00169; PH; 1.
DR   Pfam; PF00621; RhoGEF; 1.
DR   SMART; SM00233; PH; 1.
DR   SMART; SM00325; RhoGEF; 1.
DR   PROSITE; PS50003; PH_DOMAIN; 1.
SQ   SEQUENCE   790 AA;  88975 MW;  A3B9F4972CE2D509 CRC64;


  Query Match           73.9%;  Score 34;  DB 4;  Length 790;
  Best Local Similarity   62.5%;  Pred. No. 57;
  Matches    5;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 QKLCHQKK 8
          |:||||::
Db    157 QELCHQQE 164


RESULT  12
O96143
ID   O96143      PRELIMINARY;   PRT;   1188 AA.
AC   O96143;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   PROTEIN WITH 5'-3' EXONUCLEASE DOMAIN (KEM-1 FAMILY).
GN   PFB0205C.
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99021743; PubMed=9804551;
RA   Gardner M.J., Tettelin H., Carucci D.J., Cummings L.M., Aravind L.,
RA   Koonin E.V., Shallom S., Mason T., Yu K., Fujii C., Pederson J.,
RA   Shen K., Jing J., Aston C., Lai Z., Schwartz D.C., Pertea M.,
RA   Salzberg S., Zhou L., Sutton G.G., Clayton R., White O., Smith H.O.,
RA   Fraser C.M., Adams M.D., Venter J.C., Hoffman S.L.;
RT   "Chromosome 2 sequence of the human malaria parasite Plasmodium
RT   falciparum.";
RL   Science 282:1126-1132(1998).
DR   EMBL; AE001380; AAC71830.1; -.
KW   Exonuclease.
SQ   SEQUENCE   1188 AA;  142895 MW;  BF767FC8532EBAC9 CRC64;


  Query Match           73.9%;  Score 34;  DB 5;  Length 1188;
  Best Local Similarity   71.4%;  Pred. No. 76;
  Matches    5;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 QKLCHQK 7
          ::|||||
Db   1117 EELCHQK 1123


RESULT  13
O13348
ID   O13348      PRELIMINARY;      PRT;  1295 AA.
AC   O13348;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   REVERSE TRANSCRIPTASE.
OS   Magnaporthe grisea (Rice blast fungus) (Pyricularia grisea).
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC   Sordariomycetes incertae sedis; Magnaporthaceae; Magnaporthe.
OX   NCBI_TaxID=148305;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Meyn III M.A., Farrall L., Valent B., Chumley F.G., Orbach M.J.;
RL   Submitted (AUG-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: TO RNA-DIRECTED DNA POLYMERASE (REVERSE
CC       TRANSCRIPTASE).
DR   EMBL; AF018033; AAB71689.1; -.
DR   InterPro; IPR002156; RNaseH.
DR   InterPro; IPR000477; RVTse.
DR   Pfam; PF00075; rnaseH; 1.
DR   Pfam; PF00078; rvt; 1.
KW   RNA-directed DNA polymerase.
SQ   SEQUENCE   1295 AA;  144307 MW;  E811059B750D5421 CRC64;


  Query Match           73.9%;  Score 34;  DB 3;  Length 1295;
  Best Local Similarity   71.4%;  Pred. No. 81;
  Matches    5;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 KLCHQKK 8
          :||||:|
Db     33 ELCHQRK 39


RESULT  14
Q63625
ID   Q63625      PRELIMINARY;      PRT;  1473 AA.
AC   Q63625;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   CTD-BINDING SR-LIKE PROTEIN RA9.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=HIPPOCAMPUS;
RX   MEDLINE=96293459; PubMed=8692929;
RA   Yuryev A., Patturajan M., Litingtung Y., Joshi R.V., Gentile C.,
RA   Gebara M., Corden J.L.;
RT   "The C-terminal domain of the largest subunit of RNA polymerase II
RT   interacts with a novel set of serine/arginine-rich proteins.";
RL   Proc. Natl. Acad. Sci. U.S.A. 93:6975-6980(1996).
DR   EMBL; U49057; AAC52658.1; -.
SQ   SEQUENCE   1473 AA;  161203 MW;  949EE6F5873989BF CRC64;


  Query Match           73.9%;  Score 34;  DB 11;  Length 1473;
  Best Local Similarity   71.4%;  Pred. No. 89;
  Matches    5;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 QKLCHQK 7
          ||:|| |
Db   1425 QKICHSK 1431


RESULT  15
Q9P1Y6
ID   Q9P1Y6      PRELIMINARY;      PRT;  1654 AA.
AC   Q9P1Y6;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   KIAA1542 PROTEIN (FRAGMENT).
GN   KIAA1542.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20277482; PubMed=10819331;
RA   Nagase T., Kikuno R., Ishikawa K., Hirosawa M., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human
RT   genes.XVII.The complete sequences of 100 new cDNA clones from brain
```

```
RT   which code for large proteins in vitro.";
RL   DNA Res. 7:143-150(2000).
CC   -!- SIMILARITY: CONTAINS A RING-TYPE ZINC FINGER.
DR   EMBL; AB040975; BAA96066.1; -.
DR   InterPro; IPR002219; DAG_PE-bind.
DR   InterPro; IPR001965; PHD.
DR   InterPro; IPR001841; Znf_ring.
DR   Pfam; PF00628; PHD; 1.
DR   SMART; SM00109; C1; 1.
DR   SMART; SM00249; PHD; 1.
DR   SMART; SM00184; RING; 2.
DR   PROSITE; PS00518; ZINC_FINGER_C3HC4; 1.
KW   Zinc-finger.
FT   NON_TER       1      1
SQ   SEQUENCE   1654 AA;  179053 MW;  125CF71A84AFB218 CRC64;


  Query Match             73.9%;  Score 34;  DB 4;  Length 1654;
  Best Local Similarity   71.4%;  Pred. No. 97;
  Matches    5;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 QKLCHQK 7
          ||:|| |
Db   1606 QKICHSK 1612


Search completed: February 12, 2002, 12:03:26
Job time: 817 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: February 12, 2002, 12:04:02 ; Search time 30.28 Seconds
(without alignments)
9.687 Million cell updates/sec

Title: US-09-606-129A-19
Perfect score: 46
Sequence: 1 QKLCHQKK 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID		Description
1	46	100.0	295	1	BIEA_RAT	P46844	rattus norv
2	36	78.3	654	1	CBP1_YEAST	P07252	saccharomyc
3	33	71.7	673	1	Z145_HUMAN	Q05516	homo sapien
4	32	69.6	135	1	VE6_PAPVE	P11331	european el
5	32	69.6	210	1	YQS3_CAEEL	Q09311	caenorhabdi
6	32	69.6	272	1	3DHQ_ACICA	Q59087	acinetobact
7	32	69.6	345	1	TF2B_YEAST	P29055	saccharomyc
8	32	69.6	357	1	Y303_MYCGE	P47545	mycoplasma
9	32	69.6	595	1	PRIM_CHLTR	O84799	chlamydia t
10	32	69.6	609	1	FETA_GORGO	P28050	gorilla gor
11	32	69.6	609	1	FETA_HUMAN	P02771	homo sapien
12	32	69.6	645	1	YO81_CAEEL	P34617	caenorhabdi
13	32	69.6	895	1	SYI_MYCGE	P47587	mycoplasma
14	31	67.4	75	1	EX7S_BACSU	P54522	bacillus su
15	31	67.4	169	1	NEUT_BOVIN	P01156	bos taurus
16	31	67.4	170	1	NEUT_CANFA	P10673	canis famil
17	31	67.4	271	1	YSM4_CAEEL	Q10124	caenorhabdi
18	31	67.4	478	1	ARDE_CHLTR	O84375	chlamydia t
19	31	67.4	498	1	NFS1_SCHPO	O74351	schizosacch
20	31	67.4	677	1	NRG1_XENLA	O93383	xenopus lae
21	31	67.4	892	1	LDL2_XENLA	Q99088	xenopus lae
22	31	67.4	1207	1	DPOL_ASFB7	P42489	african swi
23	31	67.4	1244	1	DPOL_ASFL6	P43139	african swi
24	31	67.4	1256	1	FLII_DROME	Q24020	drosophila
25	31	67.4	1976	1	MYHA_BOVIN	Q27991	bos taurus
26	31	67.4	1976	1	MYHA_HUMAN	P35580	homo sapien
27	31	67.4	1976	1	MYHA_RAT	Q9jlt0	rattus norv
28	30	65.2	80	1	EX7S_BACHD	Q9k968	bacillus ha
29	30	65.2	159	1	MPAA_CORAV	Q08407	corylus ave
30	30	65.2	259	1	DEOC_ECOLI	P00882	escherichia
31	30	65.2	347	1	YF86_MYCPN	P75194	mycoplasma
32	30	65.2	397	1	CAPB_BACAN	P19580	bacillus an
33	30	65.2	442	1	GAG_VILV	P03352	visna lenti
34	30	65.2	442	1	GAG_VILV1	P23424	visna lenti
35	30	65.2	442	1	GAG_VILV2	P23425	visna lenti
36	30	65.2	442	1	GAG_VILVK	P35955	visna lenti
37	30	65.2	446	1	GAG_OMVVS	P16900	ovine lenti
38	30	65.2	457	1	LAT_STRCL	Q01767	streptomyce
39	30	65.2	489	1	OCLN_POTTR	Q28793	potorous tr
40	30	65.2	583	1	HASS_XENLA	O57428	xenopus lae
41	30	65.2	611	1	ACE_HAEIE	Q10715	haematobia
42	30	65.2	686	1	TRFE_ANAPL	P56410	anas platyr
43	30	65.2	705	1	TRFE_CHICK	P02789	gallus gall
44	30	65.2	753	1	YP6A_CAEEL	Q09219	caenorhabdi
45	30	65.2	837	1	SM4G_MOUSE	Q9wuh7	mus musculu

ALIGNMENTS

```
RESULT   1
BIEA_RAT
ID   BIEA_RAT       STANDARD;      PRT;   295 AA.
AC   P46844;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   BILIVERDIN REDUCTASE A PRECURSOR (EC 1.3.1.24) (BILIVERDIN-IX ALPHA-
DE   REDUCTASE).
GN   BLVRA OR BLVR.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   TISSUE=Kidney;
RX   MEDLINE=92156147; PubMed=1371282;
RA   Fakhrai H., Maines M.D.;
RT   "Expression and characterization of a cDNA for rat kidney biliverdin
RT   reductase. Evidence suggesting the liver and kidney enzymes are the
RT   same transcript product.";
RL   J. Biol. Chem. 267:4023-4029(1992).
RN   [2]
RP   MUTAGENESIS.
RX   MEDLINE=94291657; PubMed=8020496;
RA   McCoubrey W.K. Jr., Maines M.D.;
RT   "Site-directed mutagenesis of cysteine residues in biliverdin
RT   reductase. Roles in substrate and cofactor binding.";
RL   Eur. J. Biochem. 222:597-603(1994).
CC   -!- FUNCTION: CONVERTS BILIVERDIN TO BILIRUBIN; DISPLAYS TWO DISTINCT
CC       PH OPTIMA USING A DIFFERENT COFACTOR AT EACH PH: NADH AT THE LOWER
CC       PH 6.7-6.9 RANGE AND NADPH AT PH 8.5-8.7. NADPH, HOWEVER, IS THE
CC       PROBABLE COFACTOR IN BIOLOGICAL SYSTEMS.
CC   -!- CATALYTIC ACTIVITY: BILIRUBIN + NAD(P)(+) = BILIVERDIN + NAD(P)H.
CC   -!- COFACTOR: BINDS ONE ZINC ION.
CC   -!- PATHWAY: FINAL STEP IN HEME METABOLISM.
CC   -!- SUBUNIT: MONOMER (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC   -!- SIMILARITY: TO E.COLI YHHX.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M81681; AAA40830.1; -.
DR   InterPro; IPR000683; GFO_IDH_MocA.
DR   Pfam; PF01408; GFO_IDH_MocA; 1.
KW   Oxidoreductase; NAD; NADP; Zinc.
FT   PROPEP          1      2
FT   CHAIN           3    295       BILIVERDIN REDUCTASE A.
FT   DOMAIN         11     16       POLY-VAL.
```

```
FT   METAL       279    279       ZINC (POTENTIAL).
FT   METAL       280    280       ZINC (POTENTIAL).
FT   METAL       291    291       ZINC (POTENTIAL).
FT   METAL       292    292       ZINC (POTENTIAL).
FT   MUTAGEN      73     73       C->A: LOSS OF ACTIVITY.
FT   MUTAGEN     280    280       C->A: REDUCED ACTIVITY.
FT   MUTAGEN     291    291       C->A: REDUCED ACTIVITY.
SQ   SEQUENCE   295 AA;  33565 MW;  219C8EA96C150588 CRC64;


  Query Match             100.0%;  Score 46;  DB 1;  Length 295;
  Best Local Similarity   100.0%;  Pred. No. 0.18;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 QKLCHQKK 8
           ||||||||
Db     288 QKLCHQKK 295


RESULT   2
CBP1_YEAST
ID   CBP1_YEAST     STANDARD;      PRT;   654 AA.
AC   P07252;
DT   01-APR-1988 (Rel. 07, Created)
DT   01-APR-1988 (Rel. 07, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   CYTOCHROME B PRE-MRNA PROCESSING PROTEIN 1.
GN   CBP1 OR YJL209W OR J0242 OR HRA654.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=84185566; PubMed=6325407;
RA   Dieckmann C.L., Homison G., Tzagoloff A.;
RT   "Assembly of the mitochondrial membrane system. Nucleotide sequence
RT   of a yeast nuclear gene (CBP1) involved in 5' end processing of
RT   cytochrome b pre-mRNA.";
RL   J. Biol. Chem. 259:4732-4738(1984).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C;
RX   MEDLINE=95242842; PubMed=7725802;
RA   Vandenbol M., Durand P., Bolle P.-A., Dion C., Portetelle D.,
RA   Hilger F.;
RT   "Sequence analysis of a 40.2 kb DNA fragment located near the left
RT   telomere of yeast chromosome X.";
RL   Yeast 10:1657-1662(1994).
RN   [3]
RP   SEQUENCE OF 589-654 FROM N.A.
RX   MEDLINE=90014786; PubMed=2552292;
RA   Liu Y., Dieckmann C.L.;
RT   "Overproduction of yeast viruslike particles by strains deficient in
RT   a mitochondrial nuclease.";
RL   Mol. Cell. Biol. 9:3323-3331(1989).
CC   -!- FUNCTION: RESPONSIBLE FOR CONFERRING A STABLE 5'END ON CYTOCHROME
CC       B MRNA.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; K02647; AAA34474.1; -.
DR   EMBL; Z34098; CAA84002.1; -.
DR   EMBL; Z49484; CAA89506.1; -.
DR   EMBL; M28067; AAA34456.1; -.
DR   PIR; S05829; BVBYP1.
DR   PIR; S45164; S45164.
DR   SGD; S0003745; CBP1.
KW   mRNA processing.
SQ   SEQUENCE   654 AA;  76171 MW;  2453B03280E1C44D CRC64;


  Query Match             78.3%;  Score 36;  DB 1;  Length 654;
  Best Local Similarity   75.0%;  Pred. No. 21;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 QKLCHQKK 8
           :|||| ||
Db     514 KKLCHYKK 521


RESULT   3
Z145_HUMAN
ID   Z145_HUMAN     STANDARD;      PRT;   673 AA.
AC   Q05516;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   ZINC FINGER PROTEIN PLZF (PROMYELOCYTIC LEUKEMIA ZINC FINGER
DE   PROTEIN) (ZINC FINGER PROTEIN 145).
GN   ZNF145 OR PLZF.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Heart ventricle;
RX   MEDLINE=93209216; PubMed=8384553;
RA   Chen Z., Brand N.J., Chen A., Chen S.-J., Tong J.-H., Wang Z.-Y.,
RA   Waxman S., Zelent A.;
RT   "Fusion between a novel Kruppel-like zinc finger gene and the
RT   retinoic acid receptor-alpha locus due to a variant t(11;17)
RT   translocation associated with acute promyelocytic leukaemia.";
RL   EMBO J. 12:1161-1167(1993).
RN   [2]
RP   SEQUENCE OF 424-455 FROM N.A.
RX   MEDLINE=93253074; PubMed=8387545;
RA   Chen S.-J., Zelent A., Tong J.-H., Yu H.-Q., Wang Z.-Y., Derre J.,
RA   Berger R., Waxman S., Chen Z.;
RT   "Rearrangements of the retinoic acid receptor alpha and promyelocytic
RT   leukemia zinc finger genes resulting from t(11;17)(q23;q21) in a
RT   patient with acute promyelocytic leukemia.";
RL   J. Clin. Invest. 91:2260-2267(1993).
RN   [3]
RP   X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 7-122.
RX   MEDLINE=20005701; PubMed=10537309;
RA   Li X., Peng H., Schultz D.C., Lopez-Guisa J.M., Rauscher F.J. III,
RA   Marmorstein R.;
RT   "Structure-function studies of the BTB/POZ transcriptional repression
RT   domain from the promyelocytic leukemia zinc finger oncoprotein.";
RL   Cancer Res. 59:5275-5282(1999).
CC   -!- FUNCTION: PROBABLE TRANSCRIPTION FACTOR. MAY PLAY A ROLE IN
CC       MYELOID MATURATION AND IN THE DEVELOPMENT AND/OR MAINTENANCE OF
CC       OTHER DIFFERENTIATED TISSUES.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR.
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; PLFZA AND PLZFB (SHOWN HERE);
CC       ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: WITHIN THE HEMATOPOIETIC SYSTEM, PLZF IS
CC       EXPRESSED IN BONE MARROW, EARLY MYELOID CELL LINES AND PERIPHERAL
CC       BLOOD MONONUCLEAR CELLS. ALSO EXPRESSED IN THE OVARY, AND AT LOWER
CC       LEVELS, IN THE KIDNEY AND LUNG.
CC   -!- INDUCTION: BY RETINOIC ACID.
CC   -!- DISEASE: A FORM OF ACUTE PROMYELOCYTIC LEUKEMIA (APL) IS
CC       CHARACTERIZED BY A CHROMOSOMAL TRANSLOCATION T(11;17)(Q32;Q21)
CC       WHICH INVOLVES ZNF145 AND RETINOIC ACID RECEPTOR ALPHA (RARA).
CC   -!- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
CC       FINGER PROTEINS.
CC   -!- SIMILARITY: CONTAINS 1 BTB/POZ DOMAIN.
CC   ---------------------------------------------------------------------
```

```
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; Z19002; CAA79489.1; -.
DR   EMBL; S60093; AAC60590.2; -.
DR   PDB; 1CS3; 09-AUG-00.
DR   TRANSFAC; T02336; -.
DR   MIM; 176797; -.
DR   InterPro; IPR000210; BTB_POZ.
DR   InterPro; IPR000822; Znf-C2H2.
DR   Pfam; PF00651; BTB; 1.
DR   Pfam; PF00096; zf-C2H2; 9.
DR   PRINTS; PR00048; ZINCFINGER.
DR   SMART; SM00225; BTB; 1.
DR   SMART; SM00355; ZnF_C2H2; 9.
DR   PROSITE; PS50097; BTB; 1.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 8.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; 9.
KW   Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;
KW   Nuclear protein; Repeat; Chromosomal translocation; Proto-oncogene;
KW   Phosphorylation; Alternative splicing; 3D-structure.
FT   DOMAIN       34     96       BTB.
FT   DOMAIN      404    652       ZINC FINGERS.
FT   ZN_FING     404    426       C2H2-TYPE.
FT   ZN_FING     432    454       C2H2-TYPE.
FT   ZN_FING     461    483       C2H2-TYPE.
FT   ZN_FING     490    512       C2H2-TYPE.
FT   ZN_FING     518    540       C2H2-TYPE.
FT   ZN_FING     546    568       C2H2-TYPE.
FT   ZN_FING     574    596       C2H2-TYPE.
FT   ZN_FING     602    624       C2H2-TYPE.
FT   ZN_FING     630    652       C2H2-TYPE.
FT   SITE        394    395       BREAKPOINT FOR TRANSLOCATION TO FORM
FT                                PLZF-RAR-ALPHA ONCOGENE.
FT   MOD_RES      76     76       PHOSPHORYLATION (BY PDPK) (POTENTIAL).
FT   MOD_RES     184    184       PHOSPHORYLATION (BY PDPK) (POTENTIAL).
FT   MOD_RES     197    197       PHOSPHORYLATION (BY PDPK) (POTENTIAL).
FT   MOD_RES     256    256       PHOSPHORYLATION (BY PDPK) (POTENTIAL).
FT   MOD_RES     282    282       PHOSPHORYLATION (BY PDPK) (POTENTIAL).
FT   MOD_RES     628    628       PHOSPHORYLATION (BY PDPK) (POTENTIAL).
FT   VARSPLIC    255    377       MISSING (IN ISOFORM PLZFA).
SQ   SEQUENCE   673 AA;  74332 MW;  7CD7319E2A32109D CRC64;


  Query Match             71.7%;  Score 33;  DB 1;  Length 673;
  Best Local Similarity   83.3%;  Pred. No. 75;
  Matches    5;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 KLCHQK 7
          |||||:
Db    605 KLCHQR 610


RESULT   4
VE6_PAPVE
ID   VE6_PAPVE      STANDARD;      PRT;   135 AA.
AC   P11331;
DT   01-JUL-1989 (Rel. 11, Created)
DT   01-JUL-1989 (Rel. 11, Last sequence update)
DT   01-JUN-1994 (Rel. 29, Last annotation update)
DE   E6 PROTEIN.
GN   E6.
OS   European elk papillomavirus (EEPV).
OC   Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;
OC   Papillomavirus.
OX   NCBI_TaxID=10565;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87219878; PubMed=3034730;
RA   Ahola H., Bergman P., Stroem A.C., Moreno-Lopez J., Petterson U.;
RT   "Organization and expression of the transforming region from the
RT   European elk papillomavirus (EEPV).";
RL   Gene 50:195-205(1986).
CC   -!- FUNCTION: EXHIBIT A STRONG, BUT NON SPECIFIC AFFINITY FOR DOUBLE
CC       STRANDED DNA (IN VITRO).
CC   -!- SUBCELLULAR LOCATION: NUCLEAR MATRIX-ASSOCIATED.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; M15953; AAA66849.1; -.
DR   PIR; A29499; W6WLEP.
DR   InterPro; IPR001334; E6.
DR   Pfam; PF00518; E6; 1.
KW   Early protein; DNA-binding; Nuclear protein; Zinc-finger.
FT   ZN_FING      11     47       POTENTIAL.
FT   ZN_FING      83    119       POTENTIAL.
SQ   SEQUENCE   135 AA;  15869 MW;  AE4F1BABC95E0459 CRC64;


  Query Match             69.6%;  Score 32;  DB 1;  Length 135;
  Best Local Similarity   71.4%;  Pred. No. 28;
  Matches    5;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      2 KLCHQKK 8
          | ||:||
Db     24 KRCHEKK 30


RESULT   5
YQS3_CAEEL
ID   YQS3_CAEEL     STANDARD;      PRT;   210 AA.
AC   Q09311;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   HYPOTHETICAL 23.6 KDA PROTEIN F21H12.3 IN CHROMOSOME II.
GN   F21H12.3.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Favello T.;
RL   Submitted (JUL-1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: TO C.ELEGANS ZK675.1.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; U23176; AAC46715.1; -.
DR   WormPep; F21H12.3; CE01914.
KW   Hypothetical protein.
SQ   SEQUENCE   210 AA;  23617 MW;  1E646E6FC30154A0 CRC64;


  Query Match             69.6%;  Score 32;  DB 1;  Length 210;
  Best Local Similarity   71.4%;  Pred. No. 41;
  Matches    5;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;
```

```
Qy       1 QKLCHQK 7
           : |||||
Db      23 KNLCHQK 29


RESULT   6
3DHQ_ACICA
ID   3DHQ_ACICA     STANDARD;      PRT;   272 AA.
AC   Q59087;
DT   20-AUG-2001 (Rel. 40, Created)
DT   20-AUG-2001 (Rel. 40, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   CATABOLIC 3-DEHYDROQUINATE DEHYDRATASE (EC 4.2.1.10) (3-
DE   DEHYDROQUINASE).
GN   QUIB.
OS   Acinetobacter calcoaceticus.
OC   Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae;
OC   Acinetobacter.
OX   NCBI_TaxID=471;
RN   [1]
RP   SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC   STRAIN=BD413 / ADP1;
RX   MEDLINE=96011389; PubMed=7592351;
RA   Elsemore D.A., Ornston L.N.;
RT   "Unusual ancestry of dehydratases associated with quinate catabolism
RT   in Acinetobacter calcoaceticus.";
RL   J. Bacteriol. 177:5971-5978(1995).
CC   -!- FUNCTION: CATALYZES THE CATABOLIC DEHYDRATATION OF DEHYDROQUINATE
CC       TO DEHYDROSHIKIMATE.
CC   -!- CATALYTIC ACTIVITY: 3-DEHYDROQUINATE = 3-DEHYDROSHIKIMATE + H(2)O.
CC   -!- PATHWAY: QUINIC ACID CATABOLIC PATHWAY; SECOND STEP. THIS PATHWAY
CC       ALLOWS GROWTH OF BACTERIA WITH QUINATE BY ITS CONVERSION TO
CC       PROTOCATECHUATE AND SUBSEQUENT METABOLISM BY THE BETA-KETOADIPATE
CC       PATHWAY.
CC   -!- INDUCTION: BY PROTOCATECHUATE.
CC   -!- SIMILARITY: BELONGS TO THE TYPE-I 3-DEHYDROQUINASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L05770; AAC37158.1; ALT_INIT.
DR   InterPro; IPR001381; DHquinase_I.
DR   Pfam; PF01487; DHquinase_I; 1.
DR   PROSITE; PS01028; DEHYDROQUINASE_I; 1.
KW   Quinate metabolism; Lyase.
FT   ACT_SITE    163    163       BY SIMILARITY.
FT   ACT_SITE    190    190       FORMS A SCHIFF-BASE INTERMEDIATE
FT                                (BY SIMILARITY).
SQ   SEQUENCE   272 AA;  29899 MW;  98646DC5E88BF6D3 CRC64;


  Query Match             69.6%;  Score 32;  DB 1;  Length 272;
  Best Local Similarity   85.7%;  Pred. No. 51;
  Matches    6;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 KLCHQKK 8
           || ||||
Db     149 KLAHQKK 155


RESULT   7
TF2B_YEAST
ID   TF2B_YEAST     STANDARD;      PRT;   345 AA.
AC   P29055;
DT   01-DEC-1992 (Rel. 24, Created)
DT   01-DEC-1992 (Rel. 24, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   TRANSCRIPTION INITIATION FACTOR IIB (TFIIB) (TRANSCRIPTION FACTOR E).
GN   SUA7 OR YPR086W OR P9513.4.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92191276; PubMed=1547497;
RA   Pinto I., Ware D.E., Hampsey M.;
RT   "The yeast SUA7 gene encodes a homolog of human transcription factor
RT   TFIIB and is required for normal start site selection in vivo.";
RL   Cell 68:977-988(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / AB972;
RA   Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,
RA   Favello A., Fulton L., Gattung S., Greco T., Kirsten J.,
RA   Kucaba T., Hallsworth K., Hawkins J., Hillier L., Jier M.,
RA   Johnson D., Johnston L., Langston Y., Latreille P., Le T.,
RA   Mardis E., Menezes S., Miller N., Nhan M., Pauley A., Peluso D.,
RA   Rifken L., Riles L., Taich A., Trevaskis E., Vignati D.,
RA   Wilcox L., Wohldman P., Vaudin M., Wilson R., Waterston R.;
RL   Submitted (MAR-1996) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: GENERAL FACTOR THAT PLAYS A MAJOR ROLE IN THE ACTIVATION
CC       OF EUKARYOTIC GENES TRANSCRIBED BY RNA POLYMERASE II.
CC   -!- SUBUNIT: ASSOCIATES WITH TFIID-IIA (DA COMPLEX) TO FORM TFIID-
CC       IIA-IIB (DAB-COMPLEX) WHICH IS THEN RECOGNIZED BY POLYMERASE II.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR.
CC   -!- SIMILARITY: BELONGS TO THE TFIIB FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M81380; AAA35126.1; -.
DR   EMBL; U51033; AAB68135.1; -.
DR   PIR; S26707; S26707.
DR   HSSP; Q00403; 1TFB.
DR   TRANSFAC; T00819; -.
DR   SGD; S0006290; SUA7.
DR   InterPro; IPR000553; Cyclin.
DR   InterPro; IPR000812; TFIIB.
DR   Pfam; PF00382; transcript_fac2; 2.
DR   PRINTS; PR00685; TIFACTORIIB.
DR   SMART; SM00385; CYCLIN; 2.
DR   PROSITE; PS00782; TFIIB; 1.
KW   Transcription regulation; Nuclear protein; Repeat; Zinc-finger.
FT   ZN_FING      24     48       POTENTIAL.
FT   REPEAT      136    212
FT   REPEAT      242    318
SQ   SEQUENCE   345 AA;  38200 MW;  8F1F6D24602436E2 CRC64;


  Query Match             69.6%;  Score 32;  DB 1;  Length 345;
  Best Local Similarity   71.4%;  Pred. No. 63;
  Matches    5;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 KLCHQKK 8
           |||| :|
Db     155 KLCHDEK 161


RESULT   8
Y303_MYCGE
ID   Y303_MYCGE     STANDARD;      PRT;   357 AA.
AC   P47545;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
```

```
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   HYPOTHETICAL ABC TRANSPORTER ATP-BINDING PROTEIN MG303.
GN   MG303.
OS   Mycoplasma genitalium.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC   Mycoplasmataceae; Mycoplasma.
OX   NCBI_TaxID=2097;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 33530 / G-37;
RX   MEDLINE=96026346; PubMed=7569993;
RA   Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
RA   Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
RA   Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L.,
RA   Nguyen D.T., Utterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,
RA   Tomb J.-F., Dougherty B.A., Bott K.F., Hu P.-C., Lucier T.S.,
RA   Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
RT   "The minimal gene complement of Mycoplasma genitalium.";
RL   Science 270:397-403(1995).
CC   -!- SIMILARITY: BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY
CC       (ABC TRANSPORTERS).
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; U39710; AAC71525.1; -.
DR   TIGR; MG303; -.
DR   InterPro; IPR003593; AAA.
DR   InterPro; IPR003439; ABC_transportr.
DR   InterPro; IPR001687; ATP_GTP_A.
DR   Pfam; PF00005; ABC_tran; 1.
DR   SMART; SM00382; AAA; 1.
DR   PROSITE; PS00211; ABC_TRANSPORTER; FALSE_NEG.
KW   Hypothetical protein; ATP-binding; Transport; Complete proteome.
FT   NP_BIND     107    114       ATP (POTENTIAL).
SQ   SEQUENCE   357 AA;  40786 MW;  AFB1012F886E090E CRC64;


  Query Match             69.6%;  Score 32;  DB 1;  Length 357;
  Best Local Similarity   83.3%;  Pred. No. 65;
  Matches    5;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 KLCHQK 7
           ||||:|
Db     183 KLCHKK 188


RESULT   9
PRIM_CHLTR
ID   PRIM_CHLTR     STANDARD;      PRT;   595 AA.
AC   O84799;
DT   20-AUG-2001 (Rel. 40, Created)
DT   20-AUG-2001 (Rel. 40, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   DNA PRIMASE (EC 2.7.7.-).
GN   DNAG OR CT794.
OS   Chlamydia trachomatis.
OC   Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX   NCBI_TaxID=813;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=D/UW-3/CX;
RX   MEDLINE=99000809; PubMed=9784136;
RA   Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,
RA   Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
RA   Davis R.W.;
RT   "Genome sequence of an obligate intracellular pathogen of humans:
RT   Chlamydia trachomatis.";
RL   Science 282:754-759(1998).
CC   -!- FUNCTION: DNA PRIMASE IS THE POLYMERASE THAT SYNTHESIZES SMALL
CC       RNA PRIMERS FOR THE OKAZAKI FRAGMENTS ON BOTH TEMPLATE STRANDS AT
CC       REPLICATION FORKS DURING CHROMOSOMAL DNA SYNTHESIS.
CC   -!- COFACTOR: BINDS ONE ZINC ION PER MOLECULE (BY SIMILARITY).
CC   -!- SUBUNIT: MONOMER (BY SIMILARITY).
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AE001351; AAC68389.1; -.
DR   InterPro; IPR002936; Toprim.
DR   InterPro; IPR002694; ZnF_CHCC.
DR   Pfam; PF01751; Toprim; 1.
DR   Pfam; PF01807; zf-CHC2; 1.
DR   ProDom; PD002988; ZnF_CHCC; 1.
DR   SMART; SM00493; TOPRIM; 1.
DR   SMART; SM00400; ZnF_CHCC; 1.
KW   Transferase; DNA replication; DNA-directed RNA polymerase; Primosome;
KW   Zinc; Metal-binding; Complete proteome.
FT   ZN_FING     38     62       CHC2-TYPE (BY SIMILARITY).
SQ   SEQUENCE   595 AA;  68037 MW;  536858EBAFCD8FB6 CRC64;


  Query Match             69.6%;  Score 32;  DB 1;  Length 595;
  Best Local Similarity   83.3%;  Pred. No. 1e+02;
  Matches    5;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 QKLCHQ 6
             :|||||
Db     565 RKLCHQ 570


RESULT  10
FETA_GORGO
ID   FETA_GORGO     STANDARD;      PRT;   609 AA.
AC   P28050;
DT   01-AUG-1992 (Rel. 23, Created)
DT   01-AUG-1992 (Rel. 23, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   ALPHA-FETOPROTEIN PRECURSOR (ALPHA-FETOGLOBULIN) (ALPHA-1-
DE   FETOPROTEIN).
GN   AFP.
OS   Gorilla gorilla gorilla (Lowland gorilla).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Gorilla.
OX   NCBI_TaxID=9595;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91169517; PubMed=1706310;
RA   Ryan S.C., Zielinski R., Dugaiczyk A.;
RT   "Structure of the gorilla alpha-fetoprotein gene and the divergence
RT   of primates.";
RL   Genomics 9:60-72(1991).
CC   -!- FUNCTION: BINDS COPPER, NICKEL, AND FATTY ACIDS AS WELL AS, AND
CC       BILIRUBIN LESS WELL THAN, SERUM ALBUMIN.
CC   -!- SUBUNIT: DIMERIC AND TRIMERIC FORMS HAVE BEEN FOUND IN ADDITION
CC       TO THE MONOMERIC FORM (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC   -!- TISSUE SPECIFICITY: PLASMA.
CC   -!- DOMAIN: COMPOSED OF THREE HOMOLOGOUS DOMAINS.
CC   -!- PTM: SULFATED (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE ALB/AFP/VDB FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
```

```
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------------
DR   EMBL; M38272; AAA73520.1; -.
DR   PIR; A37970; FPGO.
DR   HSSP; P02768; 1UOR.
DR   InterPro; IPR000264; Serum_albumin.
DR   Pfam; PF00273; transport_prot; 3.
DR   PRINTS; PR00802; SERUMALBUMIN.
DR   PRINTS; PR00803; AFETOPROTEIN.
DR   SMART; SM00103; ALBUMIN; 3.
DR   PROSITE; PS00212; ALBUMIN; 2.
KW   Glycoprotein; Sulfation; Albumin; Plasma; Embryo; Repeat;
KW   Metal-binding; Copper; Nickel; Signal.
FT   SIGNAL       1    18       BY SIMILARITY.
FT   CHAIN       19   609       ALPHA-FETOPROTEIN.
FT   METAL       22    22       COPPER AND NICKEL (BY SIMILARITY).
FT   DISULFID    99   114       BY SIMILARITY.
FT   DISULFID   113   124       BY SIMILARITY.
FT   DISULFID   148   193       BY SIMILARITY.
FT   DISULFID   192   201       BY SIMILARITY.
FT   DISULFID   224   270       BY SIMILARITY.
FT   DISULFID   269   277       BY SIMILARITY.
FT   DISULFID   289   303       BY SIMILARITY.
FT   DISULFID   302   313       BY SIMILARITY.
FT   DISULFID   384   393       BY SIMILARITY.
FT   DISULFID   416   462       BY SIMILARITY.
FT   DISULFID   461   472       BY SIMILARITY.
FT   DISULFID   485   501       BY SIMILARITY.
FT   DISULFID   500   511       BY SIMILARITY.
FT   DISULFID   538   583       BY SIMILARITY.
FT   DISULFID   582   591       BY SIMILARITY.
FT   CARBOHYD   251   251       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   609 AA;  68697 MW;  E8AE548377DB60EB CRC64;

  Query Match             69.6%;  Score 32;  DB 1;  Length 609;
  Best Local Similarity   50.0%;  Pred. No. 1e+02;
  Matches    4;  Conservative    4;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 QKLCHQKK 8
            ::|||:|:
Db      96 EELCHEKE 103


RESULT  11
FETA_HUMAN
ID   FETA_HUMAN     STANDARD;      PRT;   609 AA.
AC   P02771;
DT   21-JUL-1986 (Rel. 01, Created)
DT   21-JUL-1986 (Rel. 01, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   ALPHA-FETOPROTEIN PRECURSOR (ALPHA-FETOGLOBULIN) (ALPHA-1-
DE   FETOPROTEIN).
GN   AFP.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=83273664; PubMed=6192439;
RA   Morinaga T., Sakai M., Wegmann T.G., Tamaoki T.;
RT   "Primary structures of human alpha-fetoprotein and its mRNA.";
RL   Proc. Natl. Acad. Sci. U.S.A. 80:4604-4608(1983).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87185438; PubMed=2436661;
RA   Gibbs P.E.M., Zielinski R., Boyd C., Dugaiczyk A.;
RT   "Structure, polymorphism, and novel repeated DNA elements revealed by
RT   a complete sequence of the human alpha-fetoprotein gene.";
RL   Biochemistry 26:1332-1343(1987).
RN   [3]
RP   SEQUENCE OF 1-28 FROM N.A.
RX   MEDLINE=93278385; PubMed=7684942;
RA   McVey J.H., Michaelides K., Hansen L.P., Ferguson-Smith M.,
RA   Tilghman S., Krumlauf R., Tuddenham E.G.D.;
RT   "A G-->A substitution in an HNF I binding site in the human alpha-
RT   fetoprotein gene is associated with hereditary persistence of alpha-
RT   fetoprotein (HPAFP).";
RL   Hum. Mol. Genet. 2:379-379(1993).
RN   [4]
RP   SEQUENCE OF 429-556 FROM N.A.
RX   MEDLINE=83158778; PubMed=6187626;
RA   Beattie W.G., Dugaiczyk A.;
RT   "Structure and evolution of human alpha-fetoprotein deduced from
RT   partial sequence of cloned cDNA.";
RL   Gene 20:415-422(1982).
RN   [5]
RP   PARTIAL SEQUENCE OF 19-609.
RX   MEDLINE=91242409; PubMed=1709810;
RA   Pucci P., Siciliano R., Malorni A., Marino G., Tecce M.F.,
RA   Ceccarini C., Terrana B.;
RT   "Human alpha-fetoprotein primary structure: a mass spectrometric
RT   study.";
RL   Biochemistry 30:5061-5066(1991).
RN   [6]
RP   PRELIMINARY SEQUENCE OF 19-35.
RX   MEDLINE=77242506; PubMed=70228;
RA   Yachnin S., Hsu R., Heinrikson R.L., Miller J.B.;
RT   "Studies on human alpha-fetoprotein. Isolation and characterization
RT   of monomeric and polymeric forms and amino-terminal sequence
RT   analysis.";
RL   Biochim. Biophys. Acta 493:418-428(1977).
RN   [7]
RP   PRELIMINARY SEQUENCE OF 19-38.
RX   MEDLINE=78001760; PubMed=71198;
RA   Aoyagi Y., Ikenaka T., Ichida F.;
RT   "Comparative chemical structures of human alpha-fetoproteins from
RT   fetal serum and from ascites fluid of a patient with hepatoma.";
RL   Cancer Res. 37:3663-3667(1977).
RN   [8]
RP   PRELIMINARY SEQUENCE OF 19-39.
RX   MEDLINE=75018719; PubMed=4138095;
RA   Ruoslahti E., Pihko H., Vaheri A., Seppala M., Virolainen M.,
RA   Konttinen A.;
RT   "Alpha fetoprotein: structure and expression in man and inbred mouse
RT   strains under normal conditions and liver injury.";
RL   Johns Hopkins Med. J. Suppl. 3:249-255(1974).
RN   [9]
RP   GENE STRUCTURE.
RX   MEDLINE=85182629; PubMed=2580830;
RA   Sakai M., Morinaga T., Urano Y., Watanabe K., Wegmann T.G.,
RA   Tamaoki T.;
RT   "The human alpha-fetoprotein gene. Sequence organization and the 5'
RT   flanking region.";
RL   J. Biol. Chem. 260:5055-5060(1985).
RN   [10]
RP   METAL-BINDING.
RX   MEDLINE=79001617; PubMed=80265;
RA   Aoyagi Y., Ikenaka T., Ichida F.;
RT   "Copper(II)-binding ability of human alpha-fetoprotein.";
RL   Cancer Res. 38:3483-3486(1978).
RN   [11]
RP   BILIRUBIN-BINDING.
RX   MEDLINE=80001710; PubMed=89900;
RA   Aoyagi Y., Ikenaka T., Ichida F.;
RT   "Alpha-Fetoprotein as a carrier protein in plasma and its bilirubin-
RT   binding ability.";
RL   Cancer Res. 39:3571-3574(1979).
RN   [12]
RP   SULFATION.
RX   PubMed=2414772;
RA   Liu M.C., Yu S., Sy J., Redman C.M., Lipmann F.;
RT   "Tyrosine sulfation of proteins from the human hepatoma cell line
```

```
RT   HepG2.";
RL   Proc. Natl. Acad. Sci. U.S.A. 82:7160-7164(1985).
CC   -!- FUNCTION: BINDS COPPER, NICKEL, AND FATTY ACIDS AS WELL AS, AND
CC       BILIRUBIN LESS WELL THAN, SERUM ALBUMIN. ONLY A SMALL PERCENTAGE
CC       (LESS THAN 2%) OF THE HUMAN AFP SHOWS ESTROGEN-BINDING PROPERTIES.
CC   -!- SUBUNIT: DIMERIC AND TRIMERIC FORMS HAVE BEEN FOUND IN ADDITION
CC       TO THE MONOMERIC FORM.
CC   -!- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC   -!- TISSUE SPECIFICITY: PLASMA. SYNTHESIZED BY THE FETAL LIVER AND
CC       YOLK SAC.
CC   -!- DEVELOPMENTAL STAGE: OCCURS IN THE PLASMA OF FETUSES MORE THAN 4
CC       WEEKS OLD, REACHES THE HIGHEST LEVELS DURING THE 12TH-16TH WEEK OF
CC       GESTATION, AND DROPS TO TRACE AMOUNTS AFTER BIRTH. THE SERUM LEVEL
CC       IN ADULTS IS USUALLY LESS THAN 40 NG/ML. AFP OCCURS ALSO AT HIGH
CC       LEVELS IN THE PLASMA AND ASCITIC FLUID OF ADULTS WITH HEPATOMA.
CC   -!- DOMAIN: COMPOSED OF THREE HOMOLOGOUS DOMAINS.
CC   -!- PTM: INDEPENDENT STUDIES SUGGEST HETEROGENEITY OF THE AMINO-
CC       TERMINAL SEQUENCE OF THE MATURE PROTEIN AND OF THE CLEAVAGE SITE
CC       OF THE SIGNAL SEQUENCE.
CC   -!- PTM: SULFATED.
CC   -!- SIMILARITY: BELONGS TO THE ALB/AFP/VDB FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M10949; AAA51674.1; -.
DR   EMBL; M10950; AAA51675.1; -.
DR   EMBL; V01514; CAA24758.1; -.
DR   EMBL; M16110; AAB58754.1; -.
DR   EMBL; J00077; AAC95396.1; -.
DR   EMBL; Z19532; CAA79592.1; -.
DR   PIR; A03234; FPHU.
DR   PIR; A26624; A26624.
DR   HSSP; P02768; 1UOR.
DR   GlycoSuiteDB; P02771; -.
DR   MIM; 104150; -.
DR   InterPro; IPR000264; Serum_albumin.
DR   Pfam; PF00273; transport_prot; 3.
DR   PRINTS; PR00802; SERUMALBUMIN.
DR   PRINTS; PR00803; AFETOPROTEIN.
DR   SMART; SM00103; ALBUMIN; 3.
DR   PROSITE; PS00212; ALBUMIN; 2.
KW   Glycoprotein; Sulfation; Albumin; Plasma; Embryo; Repeat;
KW   Metal-binding; Copper; Nickel; Signal.
FT   SIGNAL        1     18
FT   CHAIN        19    609       ALPHA-FETOPROTEIN.
FT   METAL        22     22       COPPER AND NICKEL.
FT   DISULFID     99    114
FT   DISULFID    113    124
FT   DISULFID    148    193
FT   DISULFID    192    201
FT   DISULFID    224    270
FT   DISULFID    269    277
FT   DISULFID    289    303
FT   DISULFID    302    313
FT   DISULFID    384    393
FT   DISULFID    416    462
FT   DISULFID    461    472
FT   DISULFID    485    501
FT   DISULFID    500    511
FT   DISULFID    538    583
FT   DISULFID    582    591
FT   CARBOHYD    251    251       N-LINKED (GLCNAC. . .).
FT                                /FTId=CAR_000070.
SQ   SEQUENCE   609 AA;  68677 MW;  4D4E45820E1C2D4F CRC64;

  Query Match             69.6%;  Score 32;  DB 1;  Length 609;
  Best Local Similarity   50.0%;  Pred. No. 1e+02;
  Matches    4;  Conservative    4;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 QKLCHQKK 8
            ::|||:|:
Db      96 EELCHEKE 103


RESULT  12
YO81_CAEEL
ID   YO81_CAEEL     STANDARD;      PRT;   645 AA.
AC   P34617;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   HYPOTHETICAL GTP-BINDING PROTEIN ZK1236.1 IN CHROMOSOME III.
GN   ZK1236.1.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE=94150718; PubMed=7906398;
RA   Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA   Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA   Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
RA   Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
RA   Johnston L., Jones M., Kershaw J., Kirsten J., Laisster N.,
RA   Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA   Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA   Sims M., Smaldon N., Smith A., Smith M., Sonnhammer E., Staden R.,
RA   Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
RA   Waterson R., Watson A., Weinstock L., Wilkinson-Sproat J.,
RA   Wohldman P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
CC   -!- SIMILARITY: BELONGS TO THE GTP-BINDING ELONGATION FACTOR FAMILY.
CC       LEPA SUBFAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; L13200; AAA28191.2; -.
DR   HSSP; P13551; 1ELO.
DR   WormPep; ZK1236.1; CE01446.
DR   InterPro; IPR000795; GTP_EFTU.
DR   Pfam; PF00009; GTP_EFTU; 1.
DR   PROSITE; PS00301; EFACTOR_GTP; 1.
KW   Hypothetical protein; GTP-binding.
FT   NP_BIND      49     56       GTP (POTENTIAL).
FT   NP_BIND     108    112       GTP (POTENTIAL).
FT   NP_BIND     162    165       GTP (POTENTIAL).
SQ   SEQUENCE   645 AA;  72268 MW;  3F08EA3E5FD53819 CRC64;

  Query Match             69.6%;  Score 32;  DB 1;  Length 645;
  Best Local Similarity   75.0%;  Pred. No. 1.1e+02;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 QKLCHQKK 8
            :|| ||||
Db     614 KKLSHQKK 621


RESULT  13
```

```
SYI_MYCGE
ID   SYI_MYCGE      STANDARD;      PRT;   895 AA.
AC   P47587;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   ISOLEUCYL-TRNA SYNTHETASE (EC 6.1.1.5) (ISOLEUCINE--TRNA LIGASE)
DE   (ILERS).
GN   ILES OR MG345.
OS   Mycoplasma genitalium.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC   Mycoplasmataceae; Mycoplasma.
OX   NCBI_TaxID=2097;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 33530 / G-37;
RX   MEDLINE=96026346; PubMed=7569993;
RA   Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
RA   Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
RA   Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L.,
RA   Nguyen D.T., Utterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,
RA   Tomb J.-F., Dougherty B.A., Bott K.F., Hu P.-C., Lucier T.S.,
RA   Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
RT   "The minimal gene complement of Mycoplasma genitalium.";
RL   Science 270:397-403(1995).
RN   [2]
RP   SEQUENCE OF 262-371 AND 605-711 FROM N.A.
RC   STRAIN=ATCC 33530 / G-37;
RX   MEDLINE=94075230; PubMed=8253680;
RA   Peterson S.N., Hu P.-C., Bott K.F., Hutchison C.A. III;
RT   "A survey of the Mycoplasma genitalium genome by using random
RT   sequencing.";
RL   J. Bacteriol. 175:7918-7930(1993).
CC   -!- CATALYTIC ACTIVITY: ATP + L-ISOLEUCINE + TRNA(ILE) = AMP +
CC       PYROPHOSPHATE + L-ISOLEUCYL-TRNA(ILE).
CC   -!- COFACTOR: BINDS ONE ZINC ION (BY SIMILARITY).
CC   -!- SUBUNIT: MONOMER (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC   -!- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASE FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; U39716; AAC71570.1; -.
DR   EMBL; U02196; AAD12482.1; -.
DR   EMBL; U02254; AAD12519.1; -.
DR   TIGR; MG345; -.
DR   InterPro; IPR002300; tRNA-synt_1a.
DR   InterPro; IPR001412; tRNA-synt_I.
DR   InterPro; IPR002301; tRNA-synt_ile.
DR   Pfam; PF00133; tRNA-synt_1; 1.
DR   PRINTS; PR00984; TRNASYNTHILE.
DR   PROSITE; PS00178; AA_TRNA_LIGASE_I; 1.
KW   Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW   Metal-binding; Zinc; Complete proteome.
FT   SITE         57     67       "HIGH" REGION.
FT   SITE        590    594       "KMSKS" REGION.
FT   BINDING     593    593       ATP (BY SIMILARITY).
SQ   SEQUENCE   895 AA;  104395 MW;  8C78DE6A05311B22 CRC64;


  Query Match             69.6%;  Score 32;  DB 1;  Length 895;
  Best Local Similarity   83.3%;  Pred. No. 1.4e+02;
  Matches    5;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 QKLCHQ 6
           :|||||
Db     122 RKLCHQ 127
```

```
RESULT  14
EX7S_BACSU
ID   EX7S_BACSU      STANDARD;      PRT;    75 AA.
AC   P54522;
DT   01-OCT-1996 (Rel. 34, Created)
DT   20-AUG-2001 (Rel. 40, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   PROBABLE EXODEOXYRIBONUCLEASE VII SMALL SUBUNIT (EC 3.1.11.6)
DE   (EXONUCLEASE VII SMALL SUBUNIT).
GN   XSEB.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group;
OC   Bacillus/Staphylococcus group; Bacillus.
OX   NCBI_TaxID=1423;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168 / JH642;
RA   Kobayashi Y., Mizuno M., Masuda S., Takemaru K., Hosono S.,
RA   Sato T., Takeuchi M.;
RL   Submitted (MAY-1996) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: BIDIRECTIONALLY DEGRADES SINGLE-STRANDED DNA INTO LARGE
CC       ACID-INSOLUBLE OLIGONUCLEOTIDES, WHICH ARE THEN DEGRADED FURTHER
CC       INTO SMALL ACID-SOLUBLE OLIGONUCLEOTIDES (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: EXONUCLEOLYTIC CLEAVAGE IN EITHER 5'-TO 3'-OR
CC       3'-TO 5'-DIRECTION TO YIELD 5'-PHOSPHOMONONUCLEOTIDES.
CC   -!- SUBUNIT: HETEROOLIGOMER COMPOSED OF LARGE AND SMALL SUBUNITS (BY
CC       SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE XSEB FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; D84432; BAA12574.1; ALT_INIT.
DR   EMBL; Z99116; CAB14360.1; ALT_INIT.
DR   SubtiList; BG11713; xseB.
DR   InterPro; IPR003761; Exonuc_VII_S.
DR   Pfam; PF02609; Exonuc_VII_S; 1.
KW   Hydrolase; Nuclease; Exonuclease; Complete proteome.
SQ   SEQUENCE   75 AA;  8571 MW;  6A28593DAC9F092A CRC64;


  Query Match             67.4%;  Score 31;  DB 1;  Length 75;
  Best Local Similarity   66.7%;  Pred. No. 25;
  Matches    4;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 KLCHQK 7
             |:||:|
Db      39 KMCHEK 44


RESULT  15
NEUT_BOVIN
ID   NEUT_BOVIN      STANDARD;      PRT;   169 AA.
AC   P01156;
DT   21-JUL-1986 (Rel. 01, Created)
DT   01-JUL-1989 (Rel. 11, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   NEUROTENSIN/NEUROMEDIN N PRECURSOR (NT/NMN).
GN   NTS.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
```

```
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88169625; PubMed=2832414;
RA   Kislauskis E., Bullock B., McNeil S., Dobner P.R.;
RT   "The rat gene encoding neurotensin and neuromedin N. Structure,
RT   tissue-specific expression, and evolution of exon sequences.";
RL   J. Biol. Chem. 263:4963-4968(1988).
RN   [2]
RP   SEQUENCE OF 150-162.
RC   TISSUE=Hypothalamus;
RX   MEDLINE=75095678; PubMed=1167549;
RA   Carraway R., Leeman S.E.;
RT   "The amino acid sequence of a hypothalamic peptide, neurotensin.";
RL   J. Biol. Chem. 250:1907-1911(1975).
RN   [3]
RP   SYNTHESIS OF NEUROTENSIN.
RX   MEDLINE=75095679; PubMed=1112838;
RA   Carraway R., Leeman S.E.;
RT   "The synthesis of neurotensin.";
RL   J. Biol. Chem. 250:1912-1918(1975).
CC   -!- FUNCTION: NEUROTENSIN MAY PLAY AN ENDOCRINE OR PARACRINE ROLE
CC       IN THE REGULATION OF FAT METABOLISM. IT CAUSES CONTRACTION OF
CC       SMOOTH MUSCLE.
CC   -!- SUBCELLULAR LOCATION: PACKAGED WITHIN SECRETORY VESICLES.
CC   -!- TISSUE SPECIFICITY: BRAIN AND GUT.
CC   -!- SIMILARITY: C-TERMINAL SEQUENCE SIMILARITY WITH NEUROTENSIN-
CC       RELATED PEPTIDES.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M18621; AAA30668.1; -.
DR   PIR; A01420; UNBO.
KW   Cleavage on pair of basic residues; Vasoactive; Signal.
FT   SIGNAL        1     22       POTENTIAL.
FT   CHAIN        23    147       LARGE NEUROMEDIN N (NMN-125).
FT   PEPTIDE     142    147       NEUROMEDIN N.
FT   PEPTIDE     150    162       NEUROTENSIN.
FT   MOD_RES     150    150       PYRROLIDONE CARBOXYLIC ACID.
SQ   SEQUENCE   169 AA;  19712 MW;  7B78760D5E4D7D32 CRC64;


  Query Match             67.4%;  Score 31;  DB 1;  Length 169;
  Best Local Similarity   57.1%;  Pred. No. 51;
  Matches    4;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 QKLCHQK 7
             ||:|| :
Db     107 QKICHSR 113


Search completed: February 12, 2002, 12:04:02
Job time: 798 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: February 12, 2002, 11:51:40 ; Search time 55.4 Seconds
(without alignments)
11.000 Million cell updates/sec

Title: US-09-606-129A-19
Perfect score: 46
Sequence: 1 QKLCHQKK 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_68:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	46	100.0	295	2	A42268	biliverdin reducta
2	36	78.3	654	1	BVBYP1	CBP1 protein - yea
3	34	73.9	547	2	T39478	zinc-finger protei
4	34	73.9	1188	2	A71621	protein with 5'-3'
5	34	73.9	1295	2	T30528	reverse transcript
6	34	73.9	1473	2	T31422	C-terminal domain-
7	33	71.7	263	2	E72675	hypothetical prote
8	33	71.7	673	2	S36336	probable transcrip
9	33	71.7	773	2	T00554	hypothetical prote
10	33	71.7	1247	2	T18671	hypothetical prote
11	32	69.6	135	1	W6WLEP	E6 protein - Europ
12	32	69.6	210	2	T16125	hypothetical prote
13	32	69.6	290	2	I39522	3-dehydroquinate d
14	32	69.6	345	1	S26707	transcription init
15	32	69.6	357	2	E64233	membrane transport
16	32	69.6	380	2	T24081	hypothetical prote
17	32	69.6	422	2	E96753	hypothetical prote
18	32	69.6	462	2	A46170	tektin A1 - sea ur
19	32	69.6	465	2	T27032	hypothetical prote
20	32	69.6	474	2	T00699	hypothetical prote
21	32	69.6	548	2	T22137	hypothetical prote
22	32	69.6	581	2	S44896	ZK1236.1 protein -
23	32	69.6	595	2	F71471	probable DNA prima
24	32	69.6	609	1	FPHU	alpha-fetoprotein
25	32	69.6	609	1	FPGO	alpha-fetoprotein
26	32	69.6	895	2	B64238	isoleucine--tRNA l
27	32	69.6	920	2	JC7313	aryl hydrocarbon r
28	32	69.6	1056	2	T00060	hypothetical prote
29	31	67.4	54	2	H69960	exodeoxyribonuclea
30	31	67.4	170	1	UNDG	neurotensin precur
31	31	67.4	251	2	G86368	hypothetical prote
32	31	67.4	266	2	T33411	hypothetical prote
33	31	67.4	271	2	T16421	hypothetical prote
34	31	67.4	299	2	A75401	ribulose-phosphate
35	31	67.4	364	2	T03892	hypothetical prote
36	31	67.4	436	2	D84782	probable proline t
37	31	67.4	478	2	C71523	probable shikimate
38	31	67.4	491	2	G83850	ATP-dependent DNA
39	31	67.4	498	2	T11683	probable iron-sulf
40	31	67.4	510	2	D96782	hypothetical prote
41	31	67.4	631	2	JC5803	ring finger protei
42	31	67.4	670	2	B70145	periplasmic protei
43	31	67.4	676	2	T39766	probable long-chai
44	31	67.4	704	2	T05565	hypothetical prote
45	31	67.4	977	2	S53302	H+-transporting AT

ALIGNMENTS

RESULT 1
A42268
biliverdin reductase (EC 1.3.1.24) - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 05-Nov-1999
C;Accession: A42268
R;Fakhrai, H.; Maines, M.D.
J. Biol. Chem. 267, 4023-4029, 1992
A;Title: Expression and characterization of a cDNA for rat kidney biliverdin reduc
A;Reference number: A42268; MUID:92156147
A;Accession: A42268
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: nucleic acid; protein
A;Residues: 1-295 <FAK>
A;Cross-references: GB:M81681; NID:g203177; PIDN:AAA40830.1; PID:g203178
A;Experimental source: kidney
A;Note: sequence extracted from NCBI backbone (NCBIP:82800)
C;Keywords: liver; oxidoreductase

Query Match 100.0%; Score 46; DB 2; Length 295;
Best Local Similarity 100.0%; Pred. No. 0.37;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 QKLCHQKK 8
            ||||||||
Db     288 QKLCHQKK 295
```

RESULT 2
BVBYP1
CBP1 protein - yeast (Saccharomyces cerevisiae)
N;Alternate names: protein HRA654; protein J0242; protein YJL209w
C;Species: Saccharomyces cerevisiae
C;Date: 31-Mar-1991 #sequence_revision 31-Mar-1991 #text_change 12-Nov-1999
C;Accession: S05829; S50776; S56999; S56996; S45164
R;Dieckmann, C.L.; Homison, G.; Tzagoloff, A.
J. Biol. Chem. 259, 4732-4738, 1984
A;Title: Assembly of the mitochondrial membrane system. Nucleotide sequence of a y
A;Reference number: S05829; MUID:84185566
A;Accession: S05829
A;Molecule type: DNA
A;Residues: 1-654 <DIE>
A;Cross-references: EMBL:K02647; NID:g171166; PIDN:AAA34474.1; PID:g171167
R;Vandenbol, M.; Durand, P.; Bolle, P.A.; Dion, C.; Portetelle, D.; Hilger, F.
Yeast 10, 1657-1662, 1994
A;Title: Sequence analysis of a 40.2 kb DNA fragment located near the left telomer
A;Reference number: S50701; MUID:95242842
A;Accession: S50776
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-654 <VAW>

```
A;Cross-references: EMBL:Z34098; NID:g496934; PIDN:CAA84002.1; PID:g496953
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1994
R;Vandenbol, M.; Durand, P.; Portetelle, D.; Hilger, F.
submitted to the Protein Sequence Database, September 1995
A;Reference number: S56835
A;Accession: S56999
A;Molecule type: DNA
A;Residues: 1-654 <VAN>
A;Cross-references: EMBL:Z49484; NID:g1015590; PIDN:CAA89506.1; PID:g1015591; GSPDB:GN00
R;Purnelle, B.; Coster, F.; Goffeau, A.
submitted to the Protein Sequence Database, September 1995
A;Reference number: S56977
A;Accession: S56996
A;Molecule type: DNA
A;Residues: 637-654 <PUR>
A;Cross-references: EMBL:Z49484; GSPDB:GN00010; MIPS:YJL209w
C;Genetics:
A;Gene: SGD:CBP1; MIPS:YJL209w
A;Cross-references: SGD:S0003745; MIPS:YJL209w
A;Map position: 10L
A;Genome: nuclear
C;Function:
A;Description: pre-mRNA processing
A;Note: required for correct 5' terminal processing of cytochrome b pre-mRNA
C;Superfamily: CBP1 protein
C;Keywords: mitochondrion


  Query Match             78.3%;  Score 36;  DB 1;  Length 654;
  Best Local Similarity   75.0%;  Pred. No. 44;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 QKLCHQKK 8
           :|||| ||
Db     514 KKLCHYKK 521


RESULT   3
T39478
zinc-finger protein - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 02-Sep-2000
C;Accession: T39478
R;Lyne, M.; Rajandream, M.A.; Barrell, B.G.; Lucas, M.; Gaillardin, C.
submitted to the EMBL Data Library, August 1997
A;Reference number: Z21858
A;Accession: T39478
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-547 <LYN>
A;Cross-references: EMBL:AL031349; PIDN:CAA20477.1; GSPDB:GN00067; SPDB:SPBC15D4.02
A;Experimental source: strain 972h-; cosmid c15D4
C;Genetics:
A;Gene: SPDB:SPBC15D4.02
A;Map position: 2
C;Superfamily: GAL4 zinc binuclear cluster homology
F;169-206/Domain: GAL4 zinc binuclear cluster homology <GL4>


  Query Match             73.9%;  Score 34;  DB 2;  Length 547;
  Best Local Similarity   71.4%;  Pred. No. 86;
  Matches    5;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 QKLCHQK 7
           :||||:|
Db     427 RKLCHEK 433


RESULT   4
A71621
protein with 5'-3' exonuclease domain (Kem-1 family) PFB0205c - malaria parasite (Plasmo
C;Species: Plasmodium falciparum
C;Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000
C;Accession: A71621
R;Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin,
.; Pertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith,
Science 282, 1126-1132, 1998
A;Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A;Reference number: A71600; MUID:99021743
A;Accession: A71621
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1188 <GAR>
A;Cross-references: GB:AE001380; GB:AE001362; NID:g3845120; PIDN:AAC71830.1; PID:g3
A;Experimental source: clone 3D7
C;Genetics:
A;Gene: PFB0205c


  Query Match             73.9%;  Score 34;  DB 2;  Length 1188;
  Best Local Similarity   71.4%;  Pred. No. 1.6e+02;
  Matches    5;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 QKLCHQK 7
           ::|||||
Db    1117 EELCHQK 1123


RESULT   5
T30528
reverse transcriptase - rice blast fungus
C;Species: Magnaporthe grisea (rice blast fungus)
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999
C;Accession: T30528
R;Meyn III, M.A.; Farrall, L.; Valent, B.; Chumley, F.G.; Orbach, M.J.
submitted to the EMBL Data Library, August 1997
A;Description: Magnaporthe grisea repeated DNA element MGR583 is a member of the LI
A;Reference number: Z20845
A;Accession: T30528
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1295 <MEY>
A;Cross-references: EMBL:AF018033; NID:g2454620; PID:g2454622; PIDN:AAB71689.1


  Query Match             73.9%;  Score 34;  DB 2;  Length 1295;
  Best Local Similarity   71.4%;  Pred. No. 1.7e+02;
  Matches    5;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 KLCHQKK 8
           :||||:|
Db      33 ELCHQRK 39


RESULT   6
T31422
C-terminal domain-binding protein rA9 - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 07-Dec-1999
C;Accession: T31422
R;Yuryev, A.; Patturajan, M.; Litingtung, Y.; Joshi, R.V.; Gentile, C.; Gebara, M.;
Proc. Natl. Acad. Sci. U.S.A. 93, 6975-6980, 1996
A;Title: The C-terminal domain of the largest subunit of RNA polymerase II interact
A;Reference number: Z21024; MUID:96293459
A;Accession: T31422
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1473 <YUR>
A;Cross-references: EMBL:U49057; NID:g1438533; PID:g1438534; PIDN:AAC52658.1
A;Experimental source: hippocampus


  Query Match             73.9%;  Score 34;  DB 2;  Length 1473;
  Best Local Similarity   71.4%;  Pred. No. 1.9e+02;
  Matches    5;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;
```

```
Qy       1 QKLCHQK 7
           ||:|| |
Db    1425 QKICHSK 1431
```

RESULT 7
E72675
hypothetical protein APE0826 - Aeropyrum pernix (strain K1)
C;Species: Aeropyrum pernix
C;Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000
C;Accession: E72675
R;Kawarabayasi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takah
awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; K
DNA Res. 6, 83-101, 1999
A;Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyr
A;Reference number: A72450; MUID:99310339
A;Accession: E72675
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-263 <KAW>
A;Cross-references: DDBJ:AP000060; NID:g5104188; PIDN:BAA79805.1; PID:g5104490
A;Experimental source: strain K1
C;Genetics:
A;Gene: APE0826
C;Superfamily: Aeropyrum pernix hypothetical protein APE0826

```
  Query Match             71.7%;  Score 33;  DB 2;  Length 263;
  Best Local Similarity   83.3%;  Pred. No. 73;
  Matches    5;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 LCHQKK 8
           |||||:
Db     258 LCHQKR 263
```

RESULT 8
S36336
probable transcription factor PLZF - human
C;Species: Homo sapiens (man)
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 05-Nov-1999
C;Accession: S36336; S31989
R;Chen, Z.; Brand, N.J.; Chen, A.; Chen, S.J.; Tong, J.H.; Wang, Z.Y.; Waxman, S.; Zelen
EMBO J. 12, 1161-1167, 1993
A;Title: Fusion between a novel Krueppel-like zinc finger gene and the retinoic acid rec
A;Reference number: S36336; MUID:93209216
A;Accession: S36336
A;Molecule type: mRNA
A;Residues: 1-673 <CHE>
A;Cross-references: EMBL:Z19002; NID:g38517; PIDN:CAA79489.1; PID:g38518
C;Genetics:
A;Gene: PLZF
C;Superfamily: POZ domain homology
C;Keywords: zinc finger
F;20-118/Domain: POZ domain homology <POZ>

```
  Query Match             71.7%;  Score 33;  DB 2;  Length 673;
  Best Local Similarity   83.3%;  Pred. No. 1.5e+02;
  Matches    5;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 KLCHQK 7
           |||||:
Db     605 KLCHQR 610
```

RESULT 9
T00554
hypothetical protein At2g39440 [imported] - Arabidopsis thaliana
N;Alternate names: hypothetical protein F12L6.10
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 16-Feb-2001
C;Accession: T00554; D84817
R;Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.;
submitted to the EMBL Data Library, July 1998
A;Description: Arabidopsis thaliana chromosome II BAC F12L6 genomic sequence.
A;Reference number: Z14168
A;Accession: T00554
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-773 <ROU>
A;Cross-references: EMBL:AC004218; NID:g3355463; PID:g3355473
A;Experimental source: cultivar Columbia
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallo
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Ven
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487
A;Accession: D84817
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-773 <STO>
A;Cross-references: GB:AE002093; NID:g3355473; PIDN:AAC27835.1; GSPDB:GN00139
C;Genetics:
A;Gene: At2g39440; F12L6.10
A;Map position: 2
A;Introns: 35/1; 75/1; 117/1; 159/1; 222/3; 283/3; 294/1; 506/3; 567/1

```
  Query Match             71.7%;  Score 33;  DB 2;  Length 773;
  Best Local Similarity   75.0%;  Pred. No. 1.7e+02;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 QKLCHQKK 8
           :||| |||
Db     462 KKLCDQKK 469
```

RESULT 10
T18671
hypothetical protein B0240.2 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jan-2000
C;Accession: T18671
R;White, S.
submitted to the EMBL Data Library, June 1996
A;Reference number: Z19004
A;Accession: T18671
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1247 <WIL>
A;Cross-references: EMBL:Z74026; PIDN:CAA98416.1; GSPDB:GN00023; CESP:B0240.2
A;Experimental source: clone B0240
C;Genetics:
A;Gene: CESP:B0240.2
A;Map position: 5
A;Introns: 29/2; 73/2; 108/2; 129/3; 202/2; 265/3; 401/3; 454/1; 466/3; 553/3; 594/
C;Superfamily: Caenorhabditis elegans hypothetical protein B0240.2

```
  Query Match             71.7%;  Score 33;  DB 2;  Length 1247;
  Best Local Similarity   71.4%;  Pred. No. 2.5e+02;
  Matches    5;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 KLCHQKK 8
           |:||| |
Db     265 KICHQAK 271
```

RESULT 11
W6WLEP
E6 protein - European elk papillomavirus
C;Species: European elk papillomavirus
C;Date: 31-Mar-1989 #sequence_revision 31-Mar-1989 #text_change 11-May-2000
C;Accession: A29499; F94457; F94506

R;Ahola, H.; Bergman, P.; Stroem, A.C.; Moreno-Lopez, J.; Pettersson, U.
Gene 50, 195-205, 1986
A;Title: Organization and expression of the transforming region from the European elk pa
A;Reference number: A91567; MUID:87219878
A;Accession: A29499
A;Molecule type: DNA
A;Residues: 1-135 <AHO>
A;Cross-references: GB:M15953; NID:g333025; PIDN:AAA66849.1; PID:g484015
R;Eriksson, A.
unpublished results 1987, cited by GenBank
A;Reference number: A94457
A;Accession: F94457
A;Molecule type: DNA
A;Residues: 1-135 <ERI>
A;Cross-references: GB:M15953; NID:g333025; PIDN:AAA66849.1; PID:g484015
R;Pettersson, U.
submitted to GenBank, August 1987
A;Reference number: A94506
A;Accession: F94506
A;Molecule type: DNA
A;Residues: 1-135 <PET>
A;Cross-references: GB:M15953; NID:g333025; PIDN:AAA66849.1; PID:g484015
C;Superfamily: papillomavirus E6 protein
C;Keywords: DNA binding; early protein; zinc finger

Query Match 69.6%; Score 32; DB 1; Length 135;
Best Local Similarity 71.4%; Pred. No. 65;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 KLCHQKK 8
| ||:||
Db 24 KRCHEKK 30

RESULT 12
T16125
hypothetical protein F21H12.3 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 24-Nov-1999
C;Accession: T16125
R;Favello, T.
submitted to the EMBL Data Library, July 1995
A;Description: The sequence of C. elegans cosmid F21H12.
A;Reference number: Z18464
A;Accession: T16125
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-210 <FAV>
A;Cross-references: EMBL:U23176; NID:g726404; PID:g726407; PIDN:AAC46715.1; CESP:F21H12.
A;Experimental source: strain Bristol N2
C;Genetics:
A;Gene: CESP:F21H12.3
A;Introns: 37/3; 62/2
C;Superfamily: Caenorhabditis elegans hypothetical protein F21H12.3

Query Match 69.6%; Score 32; DB 2; Length 210;
Best Local Similarity 71.4%; Pred. No. 92;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 QKLCHQK 7
: |||||
Db 23 KNLCHQK 29

RESULT 13
I39522
3-dehydroquinate dehydratase (EC 4.2.1.10), catabolic [validated] - Acinetobacter calcoa
C;Species: Acinetobacter calcoaceticus
C;Date: 19-Jul-1996 #sequence_revision 19-Jul-1996 #text_change 20-Jun-2000
C;Accession: I39522
R;Elsemore, D.A.; Ornston, L.N.
J. Bacteriol. 177, 5971-5978, 1995
A;Title: Unusual ancestry of dehydratases associated with quinate catabolism in Aci
A;Reference number: I39522; MUID:96011389
A;Accession: I39522
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-290 <RES>
A;Cross-references: EMBL:U20284; NID:g644872; PID:g644873
C;Genetics:
A;Gene: quiB
C;Function:
A;Description: EC 4.2.1.10 [validated; MUID:96011389]
C;Superfamily: 3-dehydroquinate dehydratase; 3-dehydroquinate dehydratase homology
C;Keywords: carbon-oxygen lyase; hydro-lyase
F;38-277/Domain: 3-dehydroquinate dehydratase homology <DQD>

Query Match 69.6%; Score 32; DB 2; Length 290;
Best Local Similarity 85.7%; Pred. No. 1.2e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 KLCHQKK 8
|| ||||
Db 167 KLAHQKK 173

RESULT 14
S26707
transcription initiation factor IIB - yeast (Saccharomyces cerevisiae)
N;Alternate names: protein P9513.4; protein YPR086w
C;Species: Saccharomyces cerevisiae
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
C;Accession: S26707; S69072
R;Pinto, I.; Ware, D.E.; Hampsey, M.
Cell 68, 977-988, 1992
A;Title: The yeast SUA7 gene encodes a homolog of human transcription factor TFIIB
A;Reference number: S26707; MUID:92191276
A;Accession: S26707
A;Molecule type: DNA
A;Residues: 1-345 <PIN>
A;Cross-references: EMBL:M81380; NID:g172776; PIDN:AAA35126.1; PID:g172777
A;Experimental source: strain T15
R;Couch, J.
submitted to the EMBL Data Library, March 1996
A;Description: The sequence of S. cerevisiae cosmid 9513.
A;Reference number: S69057
A;Accession: S69072
A;Molecule type: DNA
A;Residues: 1-345 <COU>
A;Cross-references: EMBL:U51033; NID:g1230676; PIDN:AAB68135.1; PID:g1230692; GSPDB
C;Genetics:
A;Gene: SGD:SUA7; MIPS:YPR086w
A;Cross-references: SGD:S0006290; MIPS:YPR086w
A;Map position: 16R
C;Superfamily: transcription initiation factor IIB; transcription initiation factor
C;Keywords: DNA binding; duplication; nucleus; transcription initiation; zinc finge
F;23-318/Domain: transcription initiation factor IIB homology <TF2B>
F;24-48/Region: zinc finger CCCC motif
F;133-210,239-313/Region: duplication

Query Match 69.6%; Score 32; DB 1; Length 345;
Best Local Similarity 71.4%; Pred. No. 1.4e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 KLCHQKK 8
|||| :|
Db 155 KLCHDEK 161

RESULT 15
E64233
membrane transport protein (glnQ) homolog MG303 - Mycoplasma genitalium

C;Species: Mycoplasma genitalium
C;Date: 17-Nov-1995 #sequence_revision 17-Nov-1995 #text_change 02-Feb-2001
C;Accession: E64233
R;Fraser, C.M.; Gocayne, J.D.; White, O.; Adams, M.D.; Clayton, R.A.; Fleischmann, R.D.; M.; Fuhrmann, J.; Nguyen, D.; Utterback, T.R.; Saudek, D.M.; Phillips, C.A.; Merrick, J. , C.A.; Venter, J.C.
Science 270, 397-403, 1995
A;Title: The minimal gene complement of Mycoplasma genitalium.
A;Reference number: A64200; MUID:96026346
A;Accession: E64233
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-357 <TIGR>
A;Cross-references: GB:U39711; GB:L43967; NID:g1045997; PID:g1046002; TIGR:MG303
A;Experimental source: strain G-37
C;Genetics:
A;Genetic code: SGC3
C;Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homology
C;Keywords: ATP; nucleotide binding; P-loop
F;90-288/Domain: ATP-binding cassette homology <ABC>
F;107-114/Region: nucleotide-binding motif A (P-loop)

Query Match 69.6%; Score 32; DB 2; Length 357;
Best Local Similarity 83.3%; Pred. No. 1.4e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```
Qy       2 KLCHQK 7
           ||||:|
Db     183 KLCHKK 188
```

Search completed: February 12, 2002, 11:51:40
Job time: 301 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: February 12, 2002, 11:49:43 ; Search time 98.92 Seconds
(without alignments)
5.991 Million cell updates/sec

Title: US-09-606-129A-19
Perfect score: 46
Sequence: 1 QKLCHQKK 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_1101:*
1: /SIDS8/gcgdata/geneseq/geneseqp/AA1980.DAT:*
2: /SIDS8/gcgdata/geneseq/geneseqp/AA1981.DAT:*
3: /SIDS8/gcgdata/geneseq/geneseqp/AA1982.DAT:*
4: /SIDS8/gcgdata/geneseq/geneseqp/AA1983.DAT:*
5: /SIDS8/gcgdata/geneseq/geneseqp/AA1984.DAT:*
6: /SIDS8/gcgdata/geneseq/geneseqp/AA1985.DAT:*
7: /SIDS8/gcgdata/geneseq/geneseqp/AA1986.DAT:*
8: /SIDS8/gcgdata/geneseq/geneseqp/AA1987.DAT:*
9: /SIDS8/gcgdata/geneseq/geneseqp/AA1988.DAT:*
10: /SIDS8/gcgdata/geneseq/geneseqp/AA1989.DAT:*
11: /SIDS8/gcgdata/geneseq/geneseqp/AA1990.DAT:*
12: /SIDS8/gcgdata/geneseq/geneseqp/AA1991.DAT:*
13: /SIDS8/gcgdata/geneseq/geneseqp/AA1992.DAT:*
14: /SIDS8/gcgdata/geneseq/geneseqp/AA1993.DAT:*
15: /SIDS8/gcgdata/geneseq/geneseqp/AA1994.DAT:*
16: /SIDS8/gcgdata/geneseq/geneseqp/AA1995.DAT:*
17: /SIDS8/gcgdata/geneseq/geneseqp/AA1996.DAT:*
18: /SIDS8/gcgdata/geneseq/geneseqp/AA1997.DAT:*
19: /SIDS8/gcgdata/geneseq/geneseqp/AA1998.DAT:*
20: /SIDS8/gcgdata/geneseq/geneseqp/AA1999.DAT:*
21: /SIDS8/gcgdata/geneseq/geneseqp/AA2000.DAT:*
22: /SIDS8/gcgdata/geneseq/geneseqp/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	36	78.3	91	22	AAB75580	Human secreted pro
2	35	76.1	551	20	AAY29955	Mouse CG1CE protei
3	34	73.9	88	21	AAB56915	Human prostate can
4	34	73.9	102	22	AAG76544	Human colon cancer
5	34	73.9	522	22	AAB95513	Human protein sequ
6	34	73.9	790	22	AAB93045	Human protein sequ
7	34	73.9	1188	21	AAB18183	Plasmodium falcipa
8	33	71.7	197	21	AAG08675	Arabidopsis thalia
9	33	71.7	242	21	AAG08674	Arabidopsis thalia
10	33	71.7	678	22	AAM25801	Human protein sequ
11	33	71.7	778	21	AAY79180	Haematopoietic ste
12	32	69.6	23	19	AAW79501	Wild-type yeast tr
13	32	69.6	23	19	AAW79502	Loss-of-function m
14	32	69.6	23	19	AAW79503	Loss-of-function m
15	32	69.6	23	19	AAW79504	Loss-of-function m
16	32	69.6	23	19	AAW79505	Loss-of-function m
17	32	69.6	23	19	AAW79506	Loss-of-function m
18	32	69.6	23	19	AAW79507	Loss-of-function m
19	32	69.6	23	19	AAW79508	Loss-of-function m
20	32	69.6	23	19	AAW79509	Loss-of-function m
21	32	69.6	23	19	AAW79510	Loss-of-function m
22	32	69.6	55	22	AAM36177	Peptide #10214 enc
23	32	69.6	88	22	AAM15135	Peptide #1569 enco
24	32	69.6	88	22	AAM27590	Peptide #1627 enco
25	32	69.6	88	22	AAM02876	Peptide #1558 enco
26	32	69.6	111	20	AAY37346	Amino acid sequenc
27	32	69.6	196	11	AAR05285	Amino acid sequenc
28	32	69.6	197	17	AAR99228	Recombinant human
29	32	69.6	197	22	AAB62068	Mature rHuAFP doma
30	32	69.6	289	22	AAU14513	Human novel protei
31	32	69.6	389	17	AAR99222	Recombinant human
32	32	69.6	389	22	AAB62071	Mature rHuAFP doma
33	32	69.6	398	20	AAW78169	Human secreted pro
34	32	69.6	404	22	AAU14277	Human novel protei
35	32	69.6	590	17	AAW01023	Alpha-foetoprotein
36	32	69.6	609	17	AAR99227	Human alpha-foetop
37	32	69.6	609	22	AAU07130	Human Alphafetopro
38	32	69.6	609	22	AAB62067	Human alpha-fetopr
39	32	69.6	920	22	AAB97388	Dioxin receptor am
40	31	67.4	53	21	AAG02555	Human secreted pro
41	31	67.4	63	22	AAU14489	Human novel protei
42	31	67.4	71	22	AAU14253	Human novel protei
43	31	67.4	80	21	AAB53346	Human colon cancer
44	31	67.4	154	19	AAW37864	Human protein comp
45	31	67.4	174	21	AAY82680	Human granulocyte

ALIGNMENTS

RESULT 1
AAB75580
ID AAB75580 standard; Protein; 91 AA.
XX
AC AAB75580;
XX
DT 06-APR-2001 (first entry)
XX
DE Human secreted protein sequence encoded by gene 28 SEQ ID NO:134.
XX
KW Human; secreted protein; immunosuppressive; antiarthritic; antirheumatic;
KW antiproliferative; cytostatic; cardiant; vasotropic; cerebroprotective;
KW nootropic; neuroprotective; antibacterial; virucide; fungicide;
KW opthalmalogical; vulnerary; autoimmune disease; cardiovascular disorder;
KW hyperproliferative disorders; cerebrovascular disorder; wound healing;
KW nervous system disorder; ocular disorder; skin ageing; chemotaxis;
KW food additive.
XX
OS Homo sapiens.
XX
PN WO200077026-A1.
XX
PD 21-DEC-2000.
XX
PF 01-JUN-2000; 2000WO-US14973.
XX
PR 11-JUN-1999; 99US-0138630.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PA (ROSE/) ROSEN C A.
XX
PI Rosen CA, Ruben SM, Komatsoulis GA;
XX
DR WPI; 2001-071258/08.

DR N-PSDB; AAF64203.
XX
PT Nucleic acid molecules encoding human secreted proteins, used in
PT preventing, treating or ameliorating a disorder, e.g. Alzheimer's and
PT Parkinson's diseases and cancers -
XX
PS Disclosure; Page 60; 542pp; English.
XX
CC Human secreted proteins AAB75506 - AAB75554 are encoded by polynucleotide
CC sequences AAF64176 - AAF64224. The specification includes amino acid
CC sequences AAB75555 - AAB75606 which represent fragments of the human
CC secreted proteins, and protein sequences with which they share homology.
CC The proteins and polynucleotides, their agonists and antagonists have
CC activities dependent on the tissues and cells in which they are
CC expressed, examples of these activities include, immunosuppressive;
CC antiarthritic; antirheumatic; antiproliferative; cytostatic; cardiant;
CC vasotropic; cerebroprotective; nootropic; neuroprotective; antibacterial;
CC virucide; fungicide; opthalmalogical; and vulnerary. The proteins,
CC polynucleotides, agonists and antagonists can be used to treat or detect
CC or diagnose various diseases and disorders including, autoimmune
CC diseases e.g. rheumatoid arthritis, hyperproliferative disorders
CC e.g. neoplasms of the breast or liver, cardiovascular disorders
CC e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,
CC angiogenesis, nervous system disorders e.g. Alzheimer's disease,
CC infections caused by bacteria, viruses and fungi and ocular disorders
CC e.g. corneal infection. The polypeptides can also be used to aid wound
CC healing and epithelial cell proliferation, to prevent skin ageing due to
CC sunburn, to maintain organs before transplantation, for supporting cell
CC culture of primary tissues, to regenerate tissues and in chemotaxis. The
CC polypeptides can also be used as a food additive or preservative to
CC increase or decrease storage capabilities. Included in the invention are
CC polynucleotide sequences AAF64167 - AAF64175 and peptide AAB75505 which
CC are used in the isolation, identification and characterisation of the
CC proteins of the invention.
XX
SQ Sequence 91 AA;

Query Match 78.3%; Score 36; DB 22; Length 91;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QKLCHQ 6
||||||
Db 30 qklchq 35

RESULT 2
AAY29955
ID AAY29955 standard; Protein; 551 AA.
XX
AC AAY29955;
XX
DT 22-NOV-1999 (first entry)
XX
DE Mouse CG1CE protein sequence.
XX
KW CG1CE; Best's macular dystrophy; mutation; diagnosis; detection;
KW BMD; age-related macular dystrophy.
XX
OS Mus sp.
XX
PN WO9943695-A1.
XX
PD 02-SEP-1999.
XX
PF 22-FEB-1999; 99WO-US03790.
XX
PR 25-FEB-1998; 98US-0075941.
PR 18-DEC-1998; 98US-0112926.
XX
PA (MERI) MERCK & CO INC.
PA (UYUP-) UNIV UPPSALA.
XX
PI Petrukhin K, Caskey CT, Metzker M, Wadelius C;
XX
DR WPI; 1999-540560/45.
DR N-PSDB; AAZ21229.
XX
PT Human and mouse polynucleotides encoding CG1CE polypeptides -
XX
PS Claim 7; Fig 8; 67pp; English.
XX
CC The present sequence represents the mouse CG1CE protein. When the CG1CE
CC gene is mutated it is responsible for Best's macular dystrophy (BMD).
CC Polynucleotides encoding CG1CE are useful for diagnosing whether a
CC patient carries a mutation in the CG1CE gene. Normal and mutated
CC CG1CE proteins are useful for identifying activators and/or inhibitors
CC of these proteins, in order to treat BMD. The CG1CE gene offers a
CC simpler and cheaper method of diagnosing BMD without the need for the
CC presence of the patient. The gene may also be useful to discovering
CC the genetic cause of age-related macular dystrophy.
XX
SQ Sequence 551 AA;

Query Match 76.1%; Score 35; DB 20; Length 551;
Best Local Similarity 62.5%; Pred. No. 1.5e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 QKLCHQKK 8
|::|| ||
Db 490 qeichmkk 497

RESULT 3
AAB56915
ID AAB56915 standard; Protein; 88 AA.
XX
AC AAB56915;
XX
DT 13-MAR-2001 (first entry)
XX
DE Human prostate cancer antigen protein sequence SEQ ID NO:1493.
XX
KW Human; prostate cancer; prostate cancer antigen; detection; diagnosis;
KW neuroprotective; cytostatic; cardioactive; immunomodulatory; muscular;
KW vulnerary; gastrointestinal; nephrotropic; antiinfective; gynaecological;
KW antibacterial; gene therapy; neural; immune; reproductive; renal;
KW gastrointestinal; pulmonary; cardiovascular; proliferative disorder;
KW wound; infectious disease.
XX
OS Homo sapiens.
XX
PN WO200055174-A1.
XX
PD 21-SEP-2000.
XX
PF 08-MAR-2000; 2000WO-US05988.
XX
PR 12-MAR-1999; 99US-0124270.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PA (ROSE/) ROSEN C A.
XX
PI Rosen CA, Ruben SM;
XX
DR WPI; 2000-587513/55.
DR N-PSDB; AAF16118.
XX
PT Prostate cancer associated gene sequences, referred to as prostate
PT cancer antigens, useful for treatment, prevention, and diagnosis of
PT disorders such as prostate cancer -
XX
PS Claim 11; Page 1930-1931; 2338pp; English.
XX

CC AAF15566 to AAF16505 encode the human prostate cancer associated
CC proteins, called prostate cancer antigens, given in AAB56363 to AAB57302.
CC The prostate cancer antigens can have neuroprotective, cytostatic,
CC cardioactive, immunomodulatory, muscular, vulnerary, gastrointestinal,
CC nephrotropic, antiinfective, gynaecological and antibacterial activities,
CC and can be used in gene therapy. The prostate cancer antigen
CC polynucleotides may be used for detection of prostate cancer, chromosome
CC identification, as chromosome markers, and for numerous other diagnostic
CC or research purposes. The prostate cancer antigens may be used to treat
CC disorders such as neural, immune, muscular, reproductive,
CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
CC disorders, wounds, and infectious diseases. AAF16506 to AAF16514 to
CC AAB57303 represent sequences used in the exemplification of the present
CC invention.
XX
SQ Sequence 88 AA;

Query Match 73.9%; Score 34; DB 21; Length 88;
Best Local Similarity 62.5%; Pred. No. 44;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QKLCHQKK 8
|:|||||::
Db 62 qelchqqe 69

RESULT 4
AAG76544
ID AAG76544 standard; Protein; 102 AA.
XX
AC AAG76544;
XX
DT 03-SEP-2001 (first entry)
XX
DE Human colon cancer antigen protein SEQ ID NO:7308.
XX
KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW colorectal carcinoma.
XX
OS Homo sapiens.
XX
PN WO200122920-A2.
XX
PD 05-APR-2001.
XX
PF 28-SEP-2000; 2000WO-US26524.
XX
PR 29-SEP-1999; 99US-0157137.
PR 03-NOV-1999; 99US-0163280.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ruben SM, Barash SC, Birse CE, Rosen CA;
XX
DR WPI; 2001-235357/24.
DR N-PSDB; AAH35949.
XX
PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT useful for preventing, diagnosing and/or treating colorectal cancers -
XX
PS Claim 11; Page 8741-8742; 9803pp; English.
XX
CC AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
CC cancer-associated nucleic acid molecules (N) and proteins (P), where
CC the proteins are collectively known as colon cancer antigens. The colon
CC cancer antigens have cytostatic activity and can be used in gene
CC therapy and vaccine production. N and P may be used in the prevention,
CC diagnosis and treatment of diseases associated with inappropriate P
CC expression. For example, N and P may be used to treat disorders
CC associated with decreased expression by rectifying mutations or deletions
CC in a patient's genome that affect the activity of P by expressing
CC inactive proteins or to supplement the patients own production of P.
CC Additionally, N may be used to produce the colon cancer-associated Ps,
CC by inserting the nucleic acids into a host cell and culturing the cell
CC to express the proteins. N and P can be used in the prevention, diagnosis
CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
CC and AAB77789 represent sequences used in the exemplification of the
CC present invention.
CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
CC missing at time of publication, meaning no sequences are present for
CC SEQ ID NO:1027 to 1052, 7921 and 7922.
XX
SQ Sequence 102 AA;

Query Match 73.9%; Score 34; DB 22; Length 102;
Best Local Similarity 83.3%; Pred. No. 50;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 3 LCHQKK 8
:|||||
Db 61 ichqkk 66

RESULT 5
AAB95513
ID AAB95513 standard; Protein; 522 AA.
XX
AC AAB95513;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human protein sequence SEQ ID NO:18081.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 8; SEQ ID 18081; 2537pp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesising 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of

CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 522 AA;

Query Match 73.9%; Score 34; DB 22; Length 522;
Best Local Similarity 62.5%; Pred. No. 2.2e+02;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QKLCHQKK 8
|:||||::
Db 157 qelchqqe 164

RESULT 6
AAB93045
ID AAB93045 standard; Protein; 790 AA.
XX
AC AAB93045;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human protein sequence SEQ ID NO:11834.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 8; SEQ ID 11834; 2537pp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesising 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 790 AA;

Query Match 73.9%; Score 34; DB 22; Length 790;
Best Local Similarity 62.5%; Pred. No. 3.2e+02;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QKLCHQKK 8
|:||||::
Db 157 qelchqqe 164

RESULT 7
AAB18183
ID AAB18183 standard; Protein; 1188 AA.
XX
AC AAB18183;
XX
DT 07-NOV-2000 (first entry)
XX
DE Plasmodium falciparum chromosome 2 related protein SEQ ID NO:40.
XX
KW Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
KW antimalarial; malaria; protozoacide; infection; insecticide.
XX
OS Plasmodium falciparum.
XX
PN WO200025728-A2.
XX
PD 11-MAY-2000.
XX
PF 05-NOV-1999; 99WO-US26796.
XX
PR 05-NOV-1998; 98US-0107131.
XX
PA (HOFF/) HOFFMAN S.
PA (CARU/) CARUCCI D.
PA (GARD/) GARDNER M.
PA (VENT/) VENTER J C.
XX
PI Hoffman S, Carucci D, Gardner M, Venter JC;
XX
DR WPI; 2000-365347/31.
XX
PT Proteins encoded by chromosome 2 of the human malarial parasite,
PT Plasmodium falciparum, useful as antimalarial vaccines and in the
PT diagnosis of P.falciparum infection -
XX
PS Disclosure; Page 101-104; 577pp; English.
XX
CC The present invention describes proteins and their fragments (I) encoded
CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
CC Also described are: (1) nucleotide sequences (II) encoding (I); and (2)
CC vaccines against P. falciparum infection comprising (I) or (II).
CC (I) and (II) are useful for the development of vaccines against
CC P. falciparum infection. (I) and polyclonal antisera or a monoclonal
CC antibody raised to immunogens comprising the sequences of (I), are
CC useful in the detection of infection with P. falciparum. Furthermore,
CC (I) (especially when they are rifins or secreted or membrane proteins)

CC can aid the identification of drugs to treat or prevent P. falciparum
CC infection, or they can be used to identify drug resistance in
CC P. falciparum. Sequencing of the Plasmodium chromosome 2 and the
CC subsequent identification of proteins encoded by it will help to expand
CC our understanding of parasite biology, a process hampered by the
CC complexity of the parasitic lifecycle, and provide new targets for
CC vaccine and drug development. Parasite resistance to drugs and mosquito
CC resistance to insecticides have led to a resurgence of malaria in many
CC parts of the world, and there is a pressing need for vaccines and new
CC drugs. AAA70078 to AAA70287 and AAB18144 to AAB18352 represent nucleotide
CC and protein sequences given in the present invention, but which are not
CC specifically mentioned within the specification.
XX
SQ Sequence 1188 AA;

Query Match 73.9%; Score 34; DB 21; Length 1188;
Best Local Similarity 71.4%; Pred. No. 4.6e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QKLCHQK 7
::|||||
Db 1117 eelchqk 1123

RESULT 8
AAG08675
ID AAG08675 standard; Protein; 197 AA.
XX
AC AAG08675;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 6307.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.

PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 71.7%; Score 33; DB 21; Length 197;
Best Local Similarity 71.4%; Pred. No. 1.4e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 KLCHQKK 8
|||| |:
Db 185 klchhkr 191

RESULT 9
AAG08674
ID AAG08674 standard; Protein; 242 AA.
XX
AC AAG08674;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 6306.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.

PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.

```
PR   29-OCT-1999;   99US-0162142.

  Query Match             71.7%;  Score 33;  DB 21;  Length 242;
  Best Local Similarity   71.4%;  Pred. No. 1.6e+02;
  Matches    5;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 KLCHQKK 8
           |||| |:
Db     230 klchhkr 236


RESULT  10
AAM25801
ID   AAM25801 standard; Protein; 678 AA.
XX
AC   AAM25801;
XX
DT   16-OCT-2001  (first entry)
XX
DE   Human protein sequence SEQ ID NO:1316.
XX
KW   Human; cancer; ulcer; HIV infection; human immunodeficiency virus;
KW   antiinflammatory; antirheumatic; antiarthritic; immunosuppressive;
KW   antibacterial; endocrine; cardiant; central nervous system; virucide;
KW   anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; anaemia;
KW   antiaggregant; haemostatic; vulnerary; antiulcer; osteopathic; eczema;
KW   dermatological; antiallergic; antiasthmatic; antidiabetic; cytostatic;
KW   neuroprotective; antidepressant; nootropic; antiparkinsonian; infection;
KW   immunostimulant; gene therapy; antisense therapy; vaccine; inflammation;
KW   antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;
KW   cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;
KW   genetic disease; haematopoietic disorder; platelet disorder; asthma;
KW   thrombocytopaenia; osteoporosis; severe combined immunodeficiency;
KW   allergic rhinitis; diabetes; multiple sclerosis; depression;
KW   Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;
KW   neurological disorder.
XX
OS   Homo sapiens.
XX
PN   WO200153455-A2.
XX
PD   26-JUL-2001.
XX
PF   22-DEC-2000; 2000WO-US35017.
XX
PR   23-DEC-1999;   99US-0471275.
PR   21-JAN-2000;  2000US-0488725.
PR   25-APR-2000;  2000US-0552317.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Drmanac RT;
XX
DR   WPI; 2001-457603/49.
DR   N-PSDB; AAH99742.
XX
PT   Isolated human polynucleotides encoding polypeptides, useful for the
PT   treatment and diagnosis of e.g. cancer, ulcers and HIV infection -
XX
PS   Claim 20; Page 274; 1217pp; English.
XX
CC   AAH99166 to AAH99904 encode the human proteins given in AAM25225 to
CC   AAM25963. The proteins can have activities based on the tissues and
CC   cells they are expressed in, such as: antiinflammatory; antirheumatic;
CC   antiarthritic; immunosuppressive; antibacterial; endocrine; cardiant;
CC   central nervous system; virucide; anti-HIV; fungicide; antimutagen;
CC   cardiovascular; antianaemic; antiaggregant; haemostatic; vulnerary;
CC   antiulcer; osteopathic; dermatological; antiallergic; antiasthmatic;
CC   antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic;
CC   antiparkinsonian; and immunostimulant. The proteins and polynucleotides
CC   encoding them can be used in gene therapy, antisense therapy and vaccine
CC   production, The proteins and polynucleotides are useful for screening for
CC   agonists or antagonists of a protein and for the treatment and diagnosis
CC   of disorders associated with the activity of a protein e.g. inflammation,
CC   rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,
CC   neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal
CC   infections, autoimmunity, genetic diseases, haematopoietic disorders,
CC   anaemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers;
CC   osteoporosis, severe combined immunodeficiency, eczema, allergic
CC   rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,
CC   Alzheimer's disease, Parkinson's disease, neurodegenerative and
CC   neurological disorders.
XX
SQ   Sequence   678 AA;


  Query Match             71.7%;  Score 33;  DB 22;  Length 678;
  Best Local Similarity   83.3%;  Pred. No. 4.1e+02;
  Matches    5;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 KLCHQK 7
           |||||:
Db     610 klchqr 615


RESULT  11
AAY79180
ID   AAY79180 standard; Protein; 778 AA.
XX
AC   AAY79180;
XX
DT   19-JUN-2000  (first entry)
XX
DE   Haematopoietic stem cell specific protein.
XX
KW   Haematopoietic stem cell; immune system disorder;
KW   leukaemia; antileukaemic; immunomodulator; therapy; mouse.
XX
OS   Mus musculus.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 9
FT                   /note= "encoded by CAS"
FT   Misc-difference 21
FT                   /note= "encoded by ASA"
FT   Misc-difference 32
FT                   /note= "encoded by SAA"
FT   Misc-difference 738
FT                   /note= "encoded by TTN"
XX
PN   WO200011168-A2.
XX
PD   02-MAR-2000.
XX
PF   20-AUG-1999;   99WO-US19052.
XX
PR   21-AUG-1998;   98US-0138132.
XX
PA   (UYPR-) UNIV PRINCETON.
XX
PI   Lemischka I,  Moore K;
XX
DR   WPI; 2000-237650/20.
DR   N-PSDB; AAZ94121.
XX
PT   Hematopoietic stem cell signaling proteins modulating replication and
PT   differentiation for treating immune system disorders and leukaemia -
XX
PS   Claim 21; Page 221-223; 256pp; English.
XX
CC   The present sequence is that of a mouse haematopoietic stem cell
CC   (HSC) specific protein.  It is an example of claimed HSC-specific
CC   proteins (see AAY79176-93) predicted from novel isolated HSC-specific
CC   nucleic acids (see AAZ94077-131).  The HSCs are especially primitive
CC   HSCs (PHSCs) such as umbilical cord cells, bone marrow cells and
CC   foetal liver cells.  The encoded proteins are growth factors,
```

CC transcription factors, splicing factors, capping factors, transport
CC proteins, translation factors or replication factors that modulate
CC HSC activity, especially differentiation or replication. The
CC invention provides claimed methods: for identifying PHSC-specific
CC nucleic acids; for generating a stem cell/progenitor cell from
CC PHSCs; for identifying the presence of a PHSC in a sample; for
CC identifying the presence in a sample of a compound that modulates
CC HSC activity; for using such a compound to treat an immune system
CC condition, especially leukaemia; for introducing exogenous nucleic
CC acid into a HSC; and for ex vivo expansion of HSCs. Also claimed
CC are vectors, host cells, and an antibody that specifically binds a
CC an HSC-specific protein.
XX
SQ Sequence 778 AA;

Query Match 71.7%; Score 33; DB 21; Length 778;
Best Local Similarity 62.5%; Pred. No. 4.7e+02;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 QKLCHQKK 8
| ||:||
Db 736 qsxchekk 743

RESULT 12
AAW79501
ID AAW79501 standard; peptide; 23 AA.
XX
AC AAW79501;
XX
DT 17-DEC-1998 (first entry)
XX
DE Wild-type yeast transcription factor IIB (region 144-160).
XX
KW Yeast transcription factor IIB; yTFIIB; cell growth; mutation; hybrid;
KW human; antifungal drug.
XX
OS Saccharomyces cerevisiae.
XX
PN WO9839355-A1.
XX
PD 11-SEP-1998.
XX
PF 13-JUN-1997; 97WO-US10404.
XX
PR 06-MAR-1997; 97US-0812175.
XX
PA (CHIL-) CHILDREN'S HOSPITAL MEDICAL CENT.
XX
PI Carson DJ, Dorsey MJ, Ma J, Shaw S, Wingfield J;
XX
DR WPI; 1998-495788/42.
XX
PT Use of peptide fragments of yeast Transcription Factor IIB - to
PT identify compounds that inhibit fungal cell growth, especially to
PT screen for antifungal drugs active against Candida albicans
XX
PS Disclosure; Figure 5; 32pp; English.
XX
CC The present sequence represents the wild-type species-specific region of
CC the yeast transcription factor IIB (yTFIIB), amino acid residues 144-166.
CC This sequence has been found to contain 4 amino acids that are vital for
CC cell growth and in vivo activity of yTFIIB, mutations of these residues
CC results in cells having severe growth defects. By the analysis of hybrid
CC molecules in yeast cells, the 4 amino acids that confer yeast specificity
CC were identified as lysine 147, cysteine 149, lysine 151, and glutamic
CC acid 152. This wild-type sequence was used to generate loss-of-function
CC mutants, by changing a number of these critical amino acids to the
CC corresponding human residues. The yTFIIB fragment containing the vital 4
CC amino acid residues provides a yeast-specific target for screening
CC libraries to identify new antifungal drugs.
XX
SQ Sequence 23 AA;

Query Match 69.6%; Score 32; DB 19; Length 23;
Best Local Similarity 71.4%; Pred. No. 29;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 KLCHQKK 8
|||| :|
Db 12 klchdek 18

RESULT 13
AAW79502
ID AAW79502 standard; peptide; 23 AA.
XX
AC AAW79502;
XX
DT 17-DEC-1998 (first entry)
XX
DE Loss-of-function mutant Yml.
XX
KW Yeast transcription factor IIB; yTFIIB; cell growth; mutation; hybrid;
KW human; antifungal drug.
XX
OS Saccharomyces cerevisiae.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Misc-difference 4
FT /note= "Wild-type Lys substituted by Val"
XX
PN WO9839355-A1.
XX
PD 11-SEP-1998.
XX
PF 13-JUN-1997; 97WO-US10404.
XX
PR 06-MAR-1997; 97US-0812175.
XX
PA (CHIL-) CHILDREN'S HOSPITAL MEDICAL CENT.
XX
PI Carson DJ, Dorsey MJ, Ma J, Shaw S, Wingfield J;
XX
DR WPI; 1998-495788/42.
XX
PT Use of peptide fragments of yeast Transcription Factor IIB - to
PT identify compounds that inhibit fungal cell growth, especially to
PT screen for antifungal drugs active against Candida albicans
XX
PS Disclosure; Figure 5; 32pp; English.
XX
CC Sequences AAW79502-W79510 are loss-of-function mutants, which were
CC generated by changing either one, three or all four of the vital 4 amino
CC acid residues of the yTFIIB species-specific wild-type sequence
CC (AAW79501) to the human equivalents. The four residues that confer
CC yeast specificity have been identified as lysine 147, cysteine 149,
CC lysine 151, and glutamic acid 152, these are vital for cell growth and in
CC vivo activity of yTFIIB, mutations of these residues results in cells
CC having severe growth defects. This particular yTFIIB derivative contains
CC a single mutation, however it was not seen to show any detectable
CC decrease in the protein's ability to support cell growth. The yTFIIB
CC fragment containing the vital 4 amino acid residues provides a
CC yeast-specific target for screening libraries to identify new antifungal
CC drugs.
XX
SQ Sequence 23 AA;

Query Match 69.6%; Score 32; DB 19; Length 23;
Best Local Similarity 71.4%; Pred. No. 29;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```
Qy      2 KLCHQKK 8
          |||| :|
Db     12 klchdek 18


RESULT  14
AAW79503
ID   AAW79503 standard; peptide; 23 AA.
XX
AC   AAW79503;
XX
DT   17-DEC-1998  (first entry)
XX
DE   Loss-of-function mutant YmC.
XX
KW   Yeast transcription factor IIB; yTFIIB; cell growth; mutation; hybrid;
KW   human; antifungal drug.
XX
OS   Saccharomyces cerevisiae.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 6
FT                   /note= "Wild-type Cys substituted by Arg"
XX
PN   WO9839355-A1.
XX
PD   11-SEP-1998.
XX
PF   13-JUN-1997;   97WO-US10404.
XX
PR   06-MAR-1997;   97US-0812175.
XX
PA   (CHIL-) CHILDREN'S HOSPITAL MEDICAL CENT.
XX
PI   Carson DJ,  Dorsey MJ,  Ma J,  Shaw S,  Wingfield J;
XX
DR   WPI; 1998-495788/42.
XX
PT   Use of peptide fragments of yeast Transcription Factor IIB - to
PT   identify compounds that inhibit fungal cell growth, especially to
PT   screen for antifungal drugs active against Candida albicans
XX
PS   Disclosure; Figure 5; 32pp; English.
XX
CC   Sequences AAW79502-W79510 are loss-of-function mutants, which were
CC   generated by changing either one, three or all four of the vital 4 amino
CC   acid residues of the yTFIIB species-specific wild-type sequence
CC   (AAW79501) to the human equivalents.  The four residues that confer
CC   yeast specificity have been identified as lysine 147, cysteine 149,
CC   lysine 151, and glutamic acid 152, these are vital for cell growth and in
CC   vivo activity of yTFIIB, mutations of these residues results in cells
CC   having severe growth defects.  This particular yTFIIB derivative contains
CC   a single mutation, however it was not seen to show any detectable
CC   decrease in the protein's ability to support cell growth.  The yTFIIB
CC   fragment containing the vital 4 amino acid residues provides a
CC   yeast-specific target for screening libraries to identify new antifungal
CC   drugs.
XX
SQ   Sequence   23 AA;


  Query Match             69.6%;  Score 32;  DB 19;  Length 23;
  Best Local Similarity   71.4%;  Pred. No. 29;
  Matches    5;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      2 KLCHQKK 8
          |||| :|
Db     12 klchdek 18


RESULT  15
AAW79504
ID   AAW79504 standard; peptide; 23 AA.
XX
AC   AAW79504;
XX
DT   17-DEC-1998  (first entry)
XX
DE   Loss-of-function mutant Ym8.
XX
KW   Yeast transcription factor IIB; yTFIIB; cell growth; mutation; hybrid;
KW   human; antifungal drug.
XX
OS   Saccharomyces cerevisiae.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 8
FT                   /note= "Wild-type Lys substituted by Asn"
XX
PN   WO9839355-A1.
XX
PD   11-SEP-1998.
XX
PF   13-JUN-1997;   97WO-US10404.
XX
PR   06-MAR-1997;   97US-0812175.
XX
PA   (CHIL-) CHILDREN'S HOSPITAL MEDICAL CENT.
XX
PI   Carson DJ,  Dorsey MJ,  Ma J,  Shaw S,  Wingfield J;
XX
DR   WPI; 1998-495788/42.
XX
PT   Use of peptide fragments of yeast Transcription Factor IIB - to
PT   identify compounds that inhibit fungal cell growth, especially to
PT   screen for antifungal drugs active against Candida albicans
XX
PS   Disclosure; Figure 5; 32pp; English.
XX
CC   Sequences AAW79502-W79510 are loss-of-function mutants, which were
CC   generated by changing either one, three or all four of the vital 4 amino
CC   acid residues of the yTFIIB species-specific wild-type sequence
CC   (AAW79501) to the human equivalents.  The four residues that confer
CC   yeast specificity have been identified as lysine 147, cysteine 149,
CC   lysine 151, and glutamic acid 152, these are vital for cell growth and in
CC   vivo activity of yTFIIB, mutations of these residues results in cells
CC   having severe growth defects.  This particular yTFIIB derivative contains
CC   a single mutation, however it was not seen to show any detectable
CC   decrease in the protein's ability to support cell growth.  The yTFIIB
CC   fragment containing the vital 4 amino acid residues provides a
CC   yeast-specific target for screening libraries to identify new antifungal
CC   drugs.
XX
SQ   Sequence   23 AA;


  Query Match             69.6%;  Score 32;  DB 19;  Length 23;
  Best Local Similarity   71.4%;  Pred. No. 29;
  Matches    5;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      2 KLCHQKK 8
          |||| :|
Db     12 klchdek 18


Search completed: February 12, 2002, 11:49:44
Job time: 500 sec
```